

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2003, 02:00:14 ; Search time 82 Seconds  
(without alignments)

12128.692 Million cell updates/sec

Title: US-09-830-762-1

Perfect score: 3243

Sequence: 1 ggagctcgtgtctcctggtg.....gtctgataaaaaaaaaa 3243

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCRTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	202.8	6.3	214	4	US-09-222-575-24
C 2	65.4	2.0	7218	1	US-08-232-463-14
C 3	44	1.4	1529	3	US-09-189-760-5
C 4	44	1.4	1529	3	US-09-188-811-5
C 5	44	1.4	1529	4	US-09-514-422-5
C 6	43.2	1.3	2494	3	US-09-189-760-1
C 7	43.2	1.3	2494	4	US-09-514-422-1
C 8	41	1.3	17656	4	US-09-433-579-3
C 9	39	1.2	650	4	US-09-328-111-189
C 10	37	1.1	2615	1	US-08-072-281-1
C 11	37	1.1	2615	1	US-08-759-446-1
C 12	37	1.1	2615	1	US-09-027-998A-1
C 13	37	1.1	3050	2	US-09-031-442A-21
C 14	37	1.1	3050	4	US-09-258-377-21
C 15	37	1.1	6265	4	US-09-129-112-3
C 16	37	1.1	11485	4	US-09-410-464-9
C 17	36.6	1.1	16535	4	US-08-961-527-74
C 18	36.4	1.1	2951	1	US-08-413-118-104
C 19	36.4	1.1	2951	3	US-08-473-446-104
C 20	36.4	1.1	3000	1	US-08-220-151-1
C 21	36.4	1.1	3000	1	US-08-413-118-1
C 22	36.4	1.1	3000	3	US-08-473-446-1
C 23	36.2	1.1	936	3	US-08-910-501-3
C 24	36.2	1.1	936	4	US-09-398-550-3
C 25	36.2	1.1	939	3	US-08-910-501-1
C 26	36.2	1.1	939	4	US-09-398-550-1
C 27	35.8	1.1	414	1	US-08-377-687-48

28	35.8	1.1	414	1	US-08-777-192-48	Sequence 48, Appl
29	35.8	1.1	414	4	US-08-971-982-48	Sequence 18, Appl
30	35.8	1.1	414	4	US-09-077-951-19	Sequence 19, Appl
C 31	35.8	1.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 32	35.6	1.1	18475	3	US-08-961-527-38	Sequence 38, Appl
C 33	35.6	1.1	51952	3	US-08-947-823-1	Sequence 1, Appl
C 34	35.4	1.1	3418	4	US-09-193-562D-29	Sequence 29, Appl
C 35	35.4	1.1	5181	4	US-08-257-573-10	Sequence 10, Appl
C 36	35.4	1.1	24183	4	US-08-943-731-3	Sequence 3, Appl
C 37	35.2	1.1	2960	3	US-08-913-842-3	Sequence 3, Appl
C 38	35.2	1.1	3231	3	US-08-632-806A-7	Sequence 7, Appl
C 39	35.2	1.1	5687	4	US-09-221-017B-368	Sequence 368, App
C 40	35.2	1.1	5852	1	US-07-867-106-2	Sequence 2, Appl
C 41	35.2	1.1	9115	4	US-07-753-520B-3	Sequence 3, Appl
C 42	35	1.1	812	4	US-09-227-357-115	Sequence 115, App
C 43	34.8	1.1	2570	2	US-09-056-075-2	Sequence 2, Appl
C 44	34.6	1.1	447	1	US-08-211-942-4	Sequence 4, Appl
C 45	34.6	1.1	732	1	US-08-211-942-6	Sequence 6, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-222-575-24/C
; Sequence 24, Application US/09222575
; Patent No. 6387697
; GENERAL INFORMATION:
; APPLICANT: Yuguo, Jiang
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
; FILE REFERENCE: 210121.470
; CURRENT APPLICATION NUMBER: US/09/08/222,575
; CURRENT FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Human
US-09-222-575-24

Query Match      6.3%; Score 202.8; DB 4; Length 214;
Best Local Similarity 96.7%; Pred. No. 1.9e-47;
Matches 207; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2593 ATGATGTCATGACAGTGTGTCACATGTCACAGCCCGTAACCTTGTATTTT 2652
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 214 AAGATGTCATGACAGTGTGTCACATGTCACAGCCCGTAACCTTGTATTTT 155

QY 2653 AGTGCATGACAGCTATACCTAGTATTTTGACATGACCCAGATATGACAGTCA 2712
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 154 AAGATGTCATGACAGCTATACCTAGTATTTTGACATGACCCAGATATGACAGTCA 95

QY 2713 CAAGATTATTCATGACAGCAATGACACCTTTGACAGCATTCCTTGAGAGGATTCGA 2772
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 94 CAAGATTATTCATGACAGCAATGACACCTTTGACAGCATTCCTTGAGAGGATTCGA 35

QY 2773 AGTGCACGCCCTGTGATTAATCTTCATCATCAT 2806
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 34 AGTGCACGCCCTGTGATTAATTTTACATCAT 1

RESULT 2
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.

```

```

1  APPLICANT: FALKNER, F. G.
2  TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
3  NUMBER OF SEQUENCES: 52
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Foley & Lardner
6  STREET: 1800 Diagonal Road, Suite 500
7  CITY: Alexandria
8  STATE: VA
9  COUNTRY: USA
10 ZIP: 22313-0299
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.25
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/232,463
20 FILING DATE:
21 CLASSIFICATION: 435
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US/07/935,313
24 FILING DATE:
25 APPLICATION NUMBER: EP 91 114 300.6
26 FILING DATE: 26-AUG-1991
27 ATTORNEY/AGENT INFORMATION:
28 NAME: BENT, Stephen A.
29 REGISTRATION NUMBER: 29,768
30 REFERENCE/DOCKET NUMBER: 30472/114 IMKU
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (703)836-9300
33 TELEFAX: (703)683-4109
34 TELEX: 899149
35
36 INFORMATION FOR SEQ ID NO: 14:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 7218 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: single
41 TOPOLOGY: linear
42 IMMEDIATE SOURCE:
43 CLONE: pTZ9pt-fl5
44
45 US-08-232-463-14

```

[illegible]

QY	2365	AAT	2367
Db	1031	ATT	1029

```

RESULT 3
US-09-189-760-5
; Sequence 5, Application US/09189760
; Patent No. 6031078
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: MNI-046CP2
; CURRENT APPLICATION NUMBER: US/09/189,760
; CURRENT FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 09/163,116
; EARLIER FILING DATE: 1998-09-29
; EARLIER APPLICATION NUMBER: 60/089,467
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: (PENDING)
; EARLIER FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(749)
US-09-189-760-5

```

	Query Match	1.48;	Score 44;	DB 3;	Length 1529;
	Best Local Similarity	48.38;	Pred. No. 0.023;		
	Matches 113;	Conservative 3;	Mismatches 118;	Indels 0;	Gaps 0;
QY	1562	TTTTGTTGATGAACATTTTAAACATTGGCCAAAGACCATGTAAACACCATTAAGTTAA	1621		
Db	873	TTTTGTGCATTTCTCTTAAAGAGAGGTGCAAAAGCTTTTGTGATGGCTGCGAGTAACTGAAA	932		
QY	1622	AATTTGTCATCTTCCACACAGATTGGAGATATATTAAGGCTTTCAGTCTGAAATGGCAA	1681		
Db	933	CAAACTCTGCATTTTGTAAAAATTARGATTTAATGGAAGCTTTAAGATTTTAAATTTGCA	992		
QY	1682	AGAGTCCCAAGATGCTGAGTTTGAACAAATTACAGTGTGCCCCAGTCAACACAGAGAGAGA	1741		
Db	993	AGGGAITCCAGAGTCTGTATTTATCTTATTTGGGGAGACACTAACMWTTCAAAGAAGCAGG	1052		
QY	1742	AAAGAGAAAATGGGCTTGAAGCTAGATCTCCCTGCATCAATCTGATGGGATTCA	1795		
Db	1053	CTGTGCAACATTTTGGTGCCAGCTAGTCTATAGATGAGATTAAACCTTTTGATTTTCA	1106		

```

? RESULT 4
? US-09-188-811-5
? Sequence 5, Application US/09188811
? Patent No. 6037148
? GENERAL INFORMATION:
? APPLICANT: Rhodadoust, Mehran
? TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
? FILE REFERENCE: NMI-046CP
? CURRENT APPLICATION NUMBER: US/09/188,811
? CURRENT FILING DATE: 1998-11-09
? EARLIER APPLICATION NUMBER: 09/163,116
? EARLIER FILING DATE: 1998-09-23
? NUMBER OF SEQ ID NOS: 8
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 5
? LENGTH: 1529
? TYPE: DNA
?

```

```

; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: (3)..(749)
;
US-09-188-811-5

```

Query Match	1.4%;	Score 44;	DB 3;	Length 1529;
Best Local Similarity	48.3%;	Pred. No. 0.023;		
Matches 113; Conservative	3;	Mismatches 118;	Indels 0;	Gaps 0;

QY	1562	TTTTGTTGTGTAAGTTTAACTTTTACCTTTGCCAAGACATCTTAAACACCAAGTTAACTGTA	1621
Db	873	TTTTGTGCATTTCTCTTAAAGAAAGTGCCTAAAGCTTTTGTGTTGCTGCAGGTAACTGAA	932
QY	1622	AATTTGTGATCTTTCCACAGATTTTGGAGATATATTAAGCTTTCAAGTTCTGTAATGGCA	1681
Db	933	CAAACTGACATTTTAAAAATATGAGTTAATGGAAGCTTTTAAGGATTTTAAATTGCA	992
QY	1682	AGAGGTCCAGATGCTGATGTTTACATTTAACAGTGTCCCCACATCAACGACAGAGGA	1741
Db	993	AGGATCCAAAGGTTCTGTATTATCTTATTGGGAGACCTTACMMTTTCAAGAAGCGGG	1052
QY	1742	AAAGAGAAATGGGCTTGAAGCTAGTATCTCCGACATCATGTGATGGGATTTCA	1795
Db	1053	CTGTGAACATTTGGGTGCCACAGTCTATTCAGATGACTTAAACCTTTGATTCTCA	1106

RESULT 5  
US-09-514-422-5

```

? APPLICANT: Rhodadoust, Mehrian
? TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
? TITLE OF INVENTION: THEEPCOR
? FILE REFERENCE: MNT-046CP2
? CURRENT APPLICATION NUMBER: US/09/514,422
? CURRENT FILING DATE: 2000-02-28
? PRIOR APPLICATION NUMBER: US/09/189,760
? PRIOR FILING DATE: 1998-11-10
? PRIOR APPLICATION NUMBER: 09/163,116
? PRIOR FILING DATE: 1998-09-29
? PRIOR APPLICATION NUMBER: 60/089,467
? PRIOR FILING DATE: 1998-06-16
? PRIOR APPLICATION NUMBER: (PENDING)
? PRIOR FILING DATE: 1998-11-09
? NUMBER OF SEQ. ID NOS: 10
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 5
? LENGTH: 1529
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (3)..(749)
? US-09-514-422-5

```

Query Match	1.4%;	Score 44;	DB 4;	Length 1529;
Best Local Similarity	48.3%;	Pred. No. 0.023;		
Matches 113; Conservative	3;	Mismatches 118;	Indels 0;	Gaps 0;

Qy	1562	TTTGTGTTGATGAAAGTTTACACTTTTCCCAAGACCATTAACACCAAGTTAACTGAA	1621
Db	873	TTTGTGCATTCCTCTAAAGAAGGTGCCAAGCTTTTGATGTGCGACGGTAACGTAA	932
Qy	1622	AATTGGATCTTCCCAACGATTTGAGCATATATAGCTTTCCAGTTCTGAAATGCA	1681
Db	933	CAAACCTGACATTTTAAAAAATTAGCATTTAATGCAAGCTTTTAAAGATTTTAAATTGCA	992
Qy	1682	AGAGGTCGAAGATGCTGATTTGACATATTACAGTGTCCCGCAGTCACCAAGAGAGAGA	1741
Db	993	AGGATCCCAAGTGTGTATTTATCTTATTGGGGAGACACTAACMNTTCAAAGAAGCAGG	1052

Oy 1742 AAGAGCAAAATGGGCTTGAACGTAGATCTCCCTGCCATCAATCGATGGCATTC 1795  
| | |||| | | | | | |  
Db 1053 CTGTGAACATTGGGTGCCACAGTGCTATCAGATGAGTTAAAACCCTTGAATCTCA 1106

RESULT 6  
US-09-189-760-1  
; Sequence 1, Application US/09189760

```

1  APPLICANT: Khodadoust, Mehran
2  TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
3  TITLE OF INVENTION: THEREFOR
4  FILE REFERENCE: MNI -046CP2
5  CURRENT APPLICATION NUMBER: US/09/189,760
6  CURRENT FILING DATE: 1998-11-10
7  EARLIER APPLICATION NUMBER: 09/163,116
8  EARLIER FILING DATE: 1998-09-29
9  EARLIER APPLICATION NUMBER: 60/089,467
10 EARLIER FILING DATE: 1998-06-16
11 EARLIER APPLICATION NUMBER: (PENDING)
12 EARLIER FILING DATE: 1998-11-09
13 NUMBER OF SEQ. ID NOS: 10
14 SOFTWARE: PatentIn Ver. 2.0
15 SEQ. ID NO. 1
16 LENGTH: 2494
17 TYPE: DNA
18 ORGANISM: Homo sapiens
19 FEATURE:
20 NAME/KEY: CDS
21 LOCATION: (164)..(1714)
22 US-09-189-760-1

```

Query Match	1.38;	Score 43.2;	DB 3;	Length 2494;
Best Local Similarity	47.4%;	Pred. No. 0.05;		
Matches 11;	Conservative 5;	Mismatches 118;	Indels 0;	Gaps 0;

QY	1562	TTTGGTTGTAAGAGTTTACGTTTGCCAAAGACATGTAAACACCAAGTAACTGTAA	1621
Db	1838	KTGTGCAATTCCTAAAGAGAGTGCCAAACCTTTTGATTGCTGCAGGTAACGTAA	1897
QY	1622	AATTTGATCTCTTCCAAACAGATTTGGAGATATATAAAGCTTTACGTTCTGAAATGGCAA	1681
Db	1898	CAACCTACAGCATTTTAAATAAATTAGATTTAATGAGAGCCTTTAAGATTTTAAATGTGA	1957
QY	1682	AGAGCTCCAGATGCTGAGTTGTAACAATTTACAGTGTCCCGCAGTCAACACAGAGAGAGA	1741
Db	1958	AGGATCCCAAGTCTCTATTATTTCTTAATTTGGGAGACACTAACMMTTCAAGAAGCAGG	2017
QY	1742	AAAGAGAAATGGGCTTGAAAGTCATGATCTTCGCCATCAATCTGATGGGATTTCA	1795
Db	2018	CTGTGAACATTTGGTGCCACAGTCTCATAGATGATTTAAACCTTTGATTCTCA	2071

RESULT 7  
US-09-514-422-1

```

1  APPLICANT: Khodadoust, Mehrian
2  TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
3  TITLE OF INVENTION: THEREFOR
4  FILE REFERENCE: MNI-046CP2
5  CURRENT APPLICATION NUMBER: US/09/514,422
6  CURRENT FILING DATE: 2000-02-28
7  PRIOR APPLICATION NUMBER: US/09/189,760
8  PRIOR FILING DATE: 1998-11-10
9  PRIOR APPLICATION NUMBER: 09/163,116
10 PRIOR FILING DATE: 1998-09-29
11 PRIOR APPLICATION NUMBER: 60/089,467
12 PRIOR FILING DATE: 1998-06-16
13 PRIOR APPLICATION NUMBER: (PENDING)
14 PRIOR FILING DATE: 1998-11-09

```

NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 1  
LENGTH: 2494  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (164)..(1714)  
US-09-514-422-1

Query Match 1.3%; Score 43.2; DB 4; Length 2494;  
Best Local Similarity 47.4%; Pred. No. 0.05;  
Matches 111; Conservative 5; Mismatches 118; Indels 0; Gaps 0;

OY 1562 TTTTGTGATGATGATTTTAACTTTGCAAGACCATGTAACACACGTTAACTGTA 1621  
DB 1838 KTTKGTGCAATTCCTTAAGAGAGTGCCTTTTGTGCTGCGAGTAAGTGA 1897  
OY 1622 AATTTGATCTTTCCACAGATTTGAGATATATAGGCTTTCAGTCTGTAATGGCA 1681  
DB 1898 CAACCTAGCATTTTAAATAATATGATTAATGAGAGCTTTAGGATTTAAATTCGA 1957  
OY 1682 AGAGTCCAAAGTGTGATTTGAACATTTAGAGTGTCCCGACGACACAGAGAGA 1741  
DB 1958 AGGATCCAAAGTGTGATTTATCTTATTTGGGAGACACTAACMTTCAAGAGCAGG 2017  
OY 1742 AAGAGAAATAGGCTTGAAGCTGATGATCTCCGACCATCATGTGATGGATTCGA 1795  
DB 2018 CTGTGACATTTGGTGGCCAGTGTATCATGATGAGTTAAACCTTTGATTCGA 2071

RESULT 8  
US-09-433-579-3  
Sequence 3, Application US/09433579  
Patent No. 644877

GENERAL INFORMATION:  
APPLICANT: Rottmann, William H.  
TITLE OF INVENTION: LSAG Gene  
FILE REFERENCE: LSAG Gene  
CURRENT APPLICATION NUMBER: US/09/433,579  
CURRENT FILING DATE: 1999-11-04  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 3  
LENGTH: 17656  
TYPE: DNA  
ORGANISM: Liquidambar styraciflua  
US-09-433-579-3

Query Match 1.3%; Score 41; DB 4; Length 17656;  
Best Local Similarity 49.8%; Pred. No. 0.6;  
Matches 104; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

OY 3035 TTGTTCTTAACTCTCTTACTAGAGAGATAATATTTTGCATTAATGAAGAAATTTT 3094  
DB 9997 TGGTACTTAACTCTCTGATATGAAATATATATATATATATATATATATATA 10056  
OY 3095 TCTAGTATATACGAGCGCTTTTATTTCTAATATGATAGATATAAATGTTAGA 3154  
DB 10057 TACTGATATATGACACGCGCTTCACATGATATTTTAAATATATATATAGGTAGA 10116  
OY 3155 TAACAGATGATTTAGATTTCCAGAGATATATTAAGTCTTAGGTATGAAGATTA 3214  
DB 10117 TAGATAGTATTTTGTATATATATAGTAGTATAGTATATACAGATATATAATNC 10176  
OY 3215 ATCATCTTGTCTGATTAATAAAAAA 3243  
DB 10177 ATAAAGATTTATATTTTATATAAAGA 10205

RESULT 9  
US-09-328-111-189/c

Sequence 189, Application US/09328111  
Patent No. 6262333  
GENERAL INFORMATION:  
APPLICANT: Endege, Wilson O.  
APPLICANT: Steinmann, Kathleen E.  
APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell, Steven E.  
APPLICANT: Carroll III, Eddie  
APPLICANT: Catino, Theodore J.  
APPLICANT: Dertl, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: Monahan, John E.  
APPLICANT: Schlegel, Robert  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
FILE REFERENCE: CCD-257 (US)  
CURRENT APPLICATION NUMBER: US/09/328,111  
CURRENT FILING DATE: 1999-06-08  
EARLIER APPLICATION NUMBER: US 60/088,801  
EARLIER FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 189  
LENGTH: 650  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)..(650)  
OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-189

Query Match 1.2%; Score 39; DB 4; Length 650;  
Best Local Similarity 49.2%; Pred. No. 0.37;  
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

OY 2947 CTTAGTCCCTTATCTGCTTTATGCTTGGGCTTGGGTAGATGATACCAATGAAC 3006  
DB 562 CCGTGTCTTTCTTCTCCCTTAATATTTGGCTTCAGATTCGCCCAATCCAGA 503  
OY 3007 ACTTCAGAGACCTCTCTCTTCCAGTGTCTTTAATCTCTTACTAGAGAGATA 3066  
DB 502 GTCTAATCTTCACCCCTTCCAGTGCAGTTAAATGACCCCTTCTAATAAGTAAG 443  
OY 3067 AATATTTGCATATATGAAGAAATTTTCTGATATATACCGACGCTTTATTTCTA 3126  
DB 442 CAAGGTTCACAGACAGAGAAGATGATTCATTTGAAAAACCATGATTTAGTTAATA 383  
OY 3127 AAA 3129  
DB 382 ATA 380

RESULT 10  
US-08-072-281-1/c  
Sequence 1, Application US/08072281  
Patent No. 5495071

GENERAL INFORMATION:  
APPLICANT: Fischhoff, David A.  
APPLICANT: Fuchs, Roy L.  
APPLICANT: Lavrik, Paul B.  
APPLICANT: McPherson, Sylvia A.  
APPLICANT: Perlak, Frederick J.  
TITLE OF INVENTION: Insect Resistant Plants  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lawrence M. Lavin, Jr., Monsanto Co., BB4F  
STREET: 700 Chesterfield Parkway No. 5495071th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: United States of America

```

;
;   ZIP: 63198
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/072,281
;   FILING DATE: 19930604
;   CLASSIFICATION: 800
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/523284
;   FILING DATE: 14-MAY-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Lavin Jr., Lawrence M.
;   REGISTRATION NUMBER: 30,768
;   REFERENCE/DOCKET NUMBER: 38-21(10629)A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (314) 537-7286
;   TELEFAX: (314) 537-6047
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2615 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 205..2139
;   US-08-072-281-1
;
Query Match      1.1%; Score 37; DB 1; Length 2615;
Best Local Similarity 54.9%; Pred. No. 2.8;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
```

```

QY 1531 ATGGAATATAAGAGAGACGAGATTGTTGATGATGATTTTAACATTGGCC 1590
    ||| || | | | | | | | | | | | | | | | | | | | | | | | |
DB 2553 ATGCAACCAATCATTTATTAAGAAATAAAGATCTTTCTATTAGCAATCAATGGGAC 2494

QY 1591 AAGACCAGTGTAAACACGAGTTAACTGTAAATTTGTGATCTTTCCACAGATTGGAG 1650
    | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2493 ATAACTCGATGGAACACAGCTCTGTATACAAATATTTGATGATCTTAAGCGCTTGAAG 2434

QY 1651 ATATATAAGCTT 1663
    ||| | | | | |
DB 2433 CTATAGAACGTTT 2421

RESULT 11
; Sequence 1, Application US/08/59446
; Patent No. 5763241
; GENERAL INFORMATION:
; APPLICANT: Fuschhoff, David A.
; APPLICANT: Fuchs, Roy L.
; APPLICANT: Lavrik, Paul B.
; APPLICANT: McPherson, Sylvia A.
; APPLICANT: Perlak, Frederick J.
; TITLE OF INVENTION: Insect Resistant Plants
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lawrence M. Lavin, Jr., Monsanto Co., BBAF
; STREET: 700 Chesterfield Parkway No. 5763241th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: United States of America
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/759,446
;   FILING DATE: 05-DEC-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/072,281
;   FILING DATE:
;   APPLICATION NUMBER: US 07/523284
;   FILING DATE: 14-MAY-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Lavin Jr., Lawrence M.
;   REGISTRATION NUMBER: 30,768
;   REFERENCE/DOCKET NUMBER: 38-21(10629)A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (314) 537-7286
;   TELEFAX: (314) 537-6047
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2615 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 205..2139
;   US-08-759-446-1
;
Query Match      1.1%; Score 37; DB 1; Length 2615;
Best Local Similarity 54.9%; Pred. No. 2.8;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
```

```

QY 1531 ATGGAATATAAGAGAGACGAGATTGTTGATGATGATTTTAACATTGGCC 1590
    ||| || | | | | | | | | | | | | | | | | | | | | | | | |
DB 2553 ATGCAACCAATCATTTATTAAGAAATAAAGATCTTTCTATTAGCAATCAATGGGAC 2494

QY 1591 AAGACCAGTGTAAACACGAGTTAACTGTAAATTTGTGATCTTTCCACAGATTGGAG 1650
    | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2493 ATAACTCGATGGAACACAGCTCTGTATACAAATATTTGATGATCTTAAGCGCTTGAAG 2434

QY 1651 ATATATAAGCTT 1663
    ||| | | | | |
DB 2433 CTATAGAACGTTT 2421

RESULT 12
; Sequence 1, Application US/09027998A
; Patent No. 6284949
; GENERAL INFORMATION:
; APPLICANT: Fuschhoff, David A
; APPLICANT: Fuchs, Roy L
; APPLICANT: Perlak, Frederick J
; TITLE OF INVENTION: Insect Resistant Plants
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White and Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,998A
; FILING DATE: 23-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L
```

; SEQ ID NO 3

```
; LENGTH: 6265
; TYPE: DNA
; ORGANISM: Dolichos biflorus
; FEATURE:
; OTHER INFORMATION: genomic sequence of NBP46 (DB46)
; NAME/KEY: exon
; LOCATION: (633)..(944)
; NAME/KEY: intron
; LOCATION: (945)..(1022)
; NAME/KEY: exon
; LOCATION: (1023)..(1151)
; NAME/KEY: intron
; LOCATION: (1152)..(1559)
; NAME/KEY: exon
; LOCATION: (1560)..(1616)
; NAME/KEY: intron
; LOCATION: (1617)..(1697)
; NAME/KEY: exon
; LOCATION: (1698)..(1790)
; US-09-129-112-3
```

## Query Match

1.1%; Score 37; DB 4; Length 6265;

Best Local Similarity 50.9%; Pred. No. 4.6;

Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
OY 3048 TCCTTACTAGAGAGATAAATATTTCATATATGAAGAAATTTTCTAGTATATAAC 3107
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6145 TACTATACAAAAATAAAAAATAAATAATTAATTAATTTCTTAATAATTAATA 6086

OY 3108 GCAGGCGTTTATTTCTAATAATGATAGATATAAATGTTAGATPACAGATGATT 3167
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6085 TTAATATTTAATAATAATAATTAATAATAATATATCAATTAATATATAATAATATT 6026

OY 3168 TTAGATTTCAGAGATATATATAAGTCTTAGCTATGAATAATAATCATC 3220
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6025 ATATATGATTAATTAATTTAATAATTTAATTAGATTCCAAAATTTATATTC 5973
```

Search completed: February 7, 2003, 05:39:02

Job time : 204 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2003, 23:46:08 ; Search time 420 Seconds  
(without alignments)  
17388.644 Million cell updates/sec

Title: US-09-830-762-1

Perfect score: 3243

Sequence: 1 gggctcgtgtctcctg99gtg.....gtctgattaaaaaaaaa 3243

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq\_101002:\*

1: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3243	100.0	3243	21	Human B-aggressive
2	2842.6	87.7	3016	24	Human Interferon-1
3	2543.8	78.4	2598	22	Human transcription
4	2533	78.1	2581	24	Human ovarian anti
5	1527.8	47.1	1695	22	Human polynucleoti
6	1526.2	47.1	1747	22	Human polynucleoti
7	1340	41.3	3024	21	Human B-aggressive
8	792	24.4	854	23	Human prostate exp
9	792	24.4	854	23	Human prostate exp

Result No.	Score	Query Match	Length	ID	Description
10	622.8	19.2	822	22	Human cervical can
11	609	18.8	820	23	Human prostate exp
12	489.4	15.1	666	22	Human cervical can
13	463.2	14.3	516	22	Human cervical can
14	449.8	13.9	481	22	Human cervical can
15	424.2	13.1	668	22	Human cervical can
16	414.8	12.8	457	22	Human cervical can
17	379.4	11.7	389	22	Human breast cance
18	362.4	11.2	411	22	Human breast cance
19	329.4	10.2	643	24	Human breast cance
20	317	9.8	342	24	Human breast cance
21	317	9.8	386	23	Human prostate exp
22	317	9.8	386	23	Human prostate exp
23	316	9.7	356	23	Human prostate exp
24	299	9.2	299	24	Human prostate exp
25	284	8.8	284	24	Human prostate exp
26	280.4	8.6	524	22	Human prostate exp
27	273	8.4	378	23	Human prostate exp
28	259	8.0	363	23	Human prostate exp
29	257	7.9	363	23	Human prostate exp
30	257	7.9	363	23	Human prostate exp
31	248.4	7.7	463	23	Human prostate exp
32	246.8	7.6	260	24	Human prostate exp
33	246.8	7.6	260	24	Human prostate exp
34	243	7.5	360	23	Human prostate exp
35	209.8	6.5	387	22	Human prostate exp
36	202.8	6.3	214	22	Human prostate exp
37	202.8	6.3	214	22	Human prostate exp
38	202.8	6.3	214	22	Human prostate exp
39	183.4	5.7	343	24	Human prostate exp
40	181.2	5.6	183	24	Human prostate exp
41	168.4	5.2	172	20	Human prostate exp
42	167.4	5.2	170	21	Human prostate exp
43	162	5.0	317	23	Human prostate exp
44	158.4	4.9	185	23	Human prostate exp
45	140.4	4.3	2269	22	Human prostate exp

#### ALIGNMENTS

RESULT 1

ID AAD00629 standard: CDNA: 3243 BP.

AC AAD00629;

29-AUG-2000 (first entry)

Human B-aggressive lymphoma (BAL) CDNA.

Human: B-aggressive lymphoma; BAL; chromosome 3q21; tumour; malignancy;

differential expression; DB-CL; Diffuse large B-cell lymphoma;

KW cytostatic; vaccine; treatment; diagnosis; non-Hodgkin's lymphoma;

KW cellular adhesion; sarcoma; carcinoma; myeloma; ss.

OS Homo sapiens.

XX

XX Key

XX 5'UTR

XX CDS

XX misc-feature

XX 3'UTR

XX misc-feature

XX

Location/Qualifiers

1..228

/\*tag= a

229..2793

/\*tag= b

/\*product= "BAL protein"

229..2790

/\*tag= c

/note= "This region is specifically claimed in Claim 1 as SEQ ID No.3"

2794..3243

/\*tag= c

278..382

/\*tag= d

/note= "Alternatively spliced sequence"

```

FT      misc_feature      2528..3132
FT      /tag="e
FT      /note="This region shows 98% identity to Soares
FT      pregnant uterus NbHPu human cDNA clone 502921"
XX
XX      WO200026231-A1.
XX
XX      11-MAY-2000.
XX
XX      29-OCT-1999; 99WO-US25439.
XX
XX      29-OCT-1998; 98US-0106383.
XX      30-OCT-1998; 98US-0106448.
XX
XX      (DAND ) DANA FARBER CANCER INST INC.
XX
XX      Shipp M, Aguilar R, Yakushijiin Y;
XX
XX      WPI; 2000-365563/31.
XX      P-PSDB; AAY71082.
XX
XX      New B-aggressive lymphoma nucleic acid for identifying cells exhibiting
XX      or predisposed to malignancies such as lymphoma, sarcoma, carcinoma and
XX      myeloma
XX
XX      Claim 1; Fig 1; 151pp; English.
PS

```

CC The present sequence is a cDNA encoding B-aggressive lymphoma (BAL).  
CC protein obtained by screening two human cDNA libraries derived from  
CC anti-immunoglobulin activated splenocytes and the Raji Burkitt's lymphoma  
CC cell line cloned into pCM8. The BAL locus is mapped to chromosome 3q21.  
CC BAL was identified to be differentially expressed in DLB-CL (diffuse  
CC large B-cell lymphoma), the most common non-Hodgkin's lymphoma, by  
CC differential display technique. It was found to be more abundant in  
CC tumours from patients with high risk fatal DLB-CL disease than low risk  
CC cured disease. BAL transcripts are most abundant in lymphoid organs e.g.  
CC spleen, foetal liver and peripheral blood, and several non-haematopoietic  
CC organs e.g. heart and skeletal muscle. BAL is involved in modulation of  
CC cellular adhesion and aggressiveness/severity of malignancy such as  
CC DLB-CL. BAL nucleic acid and protein are useful for identifying cells  
CC exhibiting or predisposed to malignancies such as lymphoma, sarcoma,  
CC carcinoma and myeloma. BAL proteins are useful as immunogens to raise  
CC anti-BAL antibodies. BAL modulators or BAL molecules can be used for  
CC prophylactic and therapeutic treatment of a subject susceptible to or  
CC having a disorder associated with aberrant BAL expression or activity,  
CC such as non-Hodgkin's lymphoma. It is also useful in vaccine preparation.  
XX  
XX Sequence 3243 BP; 990 A; 647 C; 755 G; 851 T; 0 other;

Query Match	100.0%	Score 3243;	DB 21;	Length 3243;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3243; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	GGGCTTCGCTTCCTCCGGGTGCTGACACCGTGACATCCCGCGCGCCGAGAGACTTTAGAGCTCT	60
Db	1	GGGCTTCGCTTCCTCCGGGTGCTGACACCGTGACATCCCGCGCGCCGAGAGACTTTAGAGCTCT	60
QY	61	GGAGTAGCTCTCCAGCTCTCCTCTACTCCGGGGCGCCGAGATTGTACACCCGACAGAGA	120
Db	61	GGAGTAGCTCTCCAGCTCTCCTCTACTCCGGGGCGCGAGACTTGTACACCCGACAGAGA	120
QY	121	GGGGGAGCGCGCGCGCAGAAAGTGGGCGCACCATATCTGGAAACTACAGTCTATGCTTTGA	180
Db	121	GGGGGAGCGCGCGCGCAGAAAGTGGGCGCACCATATCTGGAAACTACAGTCTATGCTTTGA	180
QY	181	AGCGCAAAAAGGAAATAAACATTTAAAGACTCCCCGGGGAGCTGGAGAGTAGGACTTTTCC	240
Db	181	AGCGCAAAAAGGAAATAAACATTTAAAGACTCCCCGGGGAGCTGGAGAGTAGGACTTTTCC	240
QY	241	ATGTGTGGCGGAGCAGCAGACTTACATGAAAAATCAGTAGATTAACCTGCTCTCACTC	300
Db	241	ATGTGTGGCGGAGCAGCAGACTTACATGAAAAATCAGTAGATTAACCTGCTCTCACTC	300

QY	301	TTTGTTCAGAAAGCTTTTGCTCAGATCTTTCTCCACTGAGAAAGGGAAATACAGAAACA	360
Db	301	TTGTTTTCAGAAAGCTTTTGCTCAGATCTTTCTCCACTGAGAAAGGGAAATACAGAAACA	360
QY	361	TGTCCTCCCTCAAGTCTCAGAGACTGAGTCTTGAGAGAAAACTATAGTTGGCAAAAT	420
Db	361	TGTCCTCCCTCAAGTCTCAGAGACTGAGTCTTGAGAGAAAACTATAGTTGGCAAAAT	420
QY	421	CCCATTTAACCAATGACTTCACAAATTTTAAAAAATATATAGCGTCAGCTGTGTGAAGTC	480
Db	421	CCCATTTAACCAATGACTTCACAAATTTTAAAAAATATATAGCGTCAGCTGTGTGAAGTC	480
QY	481	CTCCAGATATAGTTTGCGCTGATCTCTACCCCTGGTCTCTCCAGTTCAGGAAGGCAACAGC	540
Db	481	CTCCAGATATAGTTTGCGCTGATCTCTACCCCTGGTCTCTCCAGTTCAGGAAGGCAACAGC	540
QY	541	AAATCTCGCAAGTGTTCAGAAAAATCTACCTCAGATAGAGTTATTCAGTCTGAGAA	600
Db	541	AAATCTCTGCAAGTGTTCAGAAAAATCTACCTCAGATAGAGTTATTCAGTCTGAGAA	600
QY	601	GATGACCTCACACACATGCTGTGTGATGCTGTGGTGAATGACACCAGTAAAGATCTTCTG	660
Db	601	GATGACCTCACACACATGCTGTGTGATGCTGTGGTGAATGACACCAGTAAAGATCTTCTG	660
QY	661	CATGGGGAGGCGCTGGCCCTGGTAAAAGCTGTGTGATTTGTAATCCAAAGAG	720
Db	661	CATGGGGAGGCGCTGGCCCTGGTAAAAGCTGTGTGATTTGTAATCCAAAGAG	720
QY	721	AGCAAAAGCTTTGTGTGGCAGATATAGTAAAGTGTACAGCTGAGAGATACCTCACGGGA	780
Db	721	AGCAAAAGCTTTGTGTGGCAGATATAGTAAAGTGTACAGCTGAGAGATACCTCACGGGA	780
QY	781	GCAGGAGGCTTCCCTGCACAAACAGATCATGCTGTGTGGCGCTCGGTGATGGAATGG	840
Db	781	GCAGGAGGCTTCCCTGCACAAACAGATCATGCTGTGTGGCGCTCGGTGATGGAATGG	840
QY	841	GATAAACAAGGATGATGCTGAGAAAGCTCAGAAAGGCCATTTGTAAATATATATGTC	900
Db	841	GATAAACAAGGATGATGCTGAGAAAGCTCAGAAAGGCCATTTGTAAATATATGTC	900
QY	901	ATCTATTAATAATTACTCACAATTAAAGACAGTATGCAATTTCCAGCCTTGAGCTGGAAATTTT	960
Db	901	ATCTATTAATAATTACTCACAATTAAAGACAGTATGCAATTTCCAGCCTTGAGCTGGAAATTTT	960
QY	961	CAGTCCCTCTGAATTTGTGTACAAAGACTATTTGTAGAGACTATCCGGGGTTAGTTTGCAA	1020
Db	961	CAGTCCCTCTGAATTTGTGTACAAAGACTATTTGTAGAGACTATCCGGGGTTAGTTTGCAA	1020
QY	1021	GGGAAGCCATGATGATGATTAATTGTAAGAAATTCACCTGGTAGCAATATAGAACCCCTACT	1080
Db	1021	GGGAAGCCATGATGATGATTAATTGTAAGAAATTCACCTGGTAGCAATATAGAACCCCTACT	1080
QY	1081	GTTGCTGCTTTAAAGAGTGGCTTCGGAATTCCTATGGGAAGATAGCTGGGCAACAACA	1140
Db	1081	GTTGCTGCTTTAAAGAGTGGCTTCGGAATTCCTATGGGAAGATAGCTGGGCAACAACA	1140
QY	1141	ACCACCCCTTTCTCAATGCAATGCAATGTGTCGAMCAACCTGACCCTCCAGATTGTCCAGGGC	1200
Db	1141	ACCACCCCTTTCTCAATGCAATGCAATGTGTCGAMCAACCTGACCCTCCAGATTGTCCAGGGC	1200
QY	1201	CACATTTGATGCGACAGCGCAGATGTATTTGTTAAATCTGTATTAACCCACATGATATTACA	1260
Db	1201	CACATTTGATGCGACAGCGCAGATGTATTTGTTAAATCTGTATTAACCCACATGATATTACA	1260
QY	1261	GTTGGACCTGTGGCAAGTCAATTTCTATCAACAAGCAGAGTTGMAATGAAATCGGAATTT	1320
Db	1261	GTTGGACCTGTGGCAAGTCAATTTCTATCAACAAGCAGAGTTGMAATGAAATCGGAATTT	1320
QY	1321	CTTGCCACAAGGCTTAAACAGTTTCAACGGCTCCCACTGTGTGCTGCACAAAGAGATTT	1380
Db	1321	CTTGCCACAAGGCTTAAACAGTTTCAACGGCTCCCACTGTGTGCTGCACAAAGAGATTT	1380
QY	1381	AACTTGTTCTGTAATATATATATACATGTACTGTGGCATTTCAAGATTTCTTAACCTCAG	1440

QY	1081	GTTGGCTGCTTAAAGCTGCCTCAGAATTCCTTAGGGAAGATGAGTGCGGGACAACAA	1140
Db	1081	GTTGGCTGCTTAAAGCTGCCTCAGAATTCCTTAGGGAAGATGAGTGCGGGACAACAA	1140
QY	1141	ACCACCCCTTTTTCATGCAATGTCGTGAACAACCTGACCCCTCCAGATTGTCACGAGC	1200
Db	1141	ACCACCCCTTTTTCATGCAATGTCGTGAACAACCTGACCCCTCCAGATTGTCACGAGC	1200
QY	1201	CACATTGTAATGGCAGACGGCAGATGTAAATTCTGTAAACCCACATGATATTACA	1260
Db	1201	CACATTGTAATGGCAGACGGCAGATGTAAATTCTGTAAACCCACATGATATTACA	1260
QY	1261	GTTGGACTGTGGCAAAGTCAATTCCTAACCAAGCAGAGATTGAATAACCGSAATTT	1320
Db	1261	GTTGGACTGTGGCAAAGTCAATTCCTAACCAAGCAGAGATTGAATAACCGSAATTT	1320
QY	1321	CTTGCCACAAGAAGGCTTAAACAAGATTTCACACGGTCCACTGTGGTACTGGTTCACAAAGATT	1380
Db	1321	CTTGCCACAAGAAGGCTTAAACAAGATTTCACACGGTCCACTGTGGTACTGGTTCACAAAGATT	1380
QY	1381	AAC TTG TCT G T A A A T T A T A T A C C A T G T A C T G T G C A T T C A G A A T T T C T A A A C C T C A G	1440
Db	1381	AAC TTG TCT G T A A A T T A T A T A C C A T G T A C T G T G C A T T C A G A A T T T C T A A A C C T C A G	1440

```

|||||
Db 1381 AACTGTTCTGTAATATATATACCATCTACTGTGTCATTCAGAAATTTCTCTAAACCTCAG 1440
|||
Qy 1441 AATATTAACATGCATGAGAGAGTGTGGAAAAATGCATTTAGAGCAAAATATACCTCC 1500
|||
Db 1441 AATATTAACATGCATGAGAGAGTGTGGAAAAATGCATTTAGAGCAAAATATACCTCC 1500
|||
Qy 1501 AATTCCTTCTGCTCCCTTGGAGCTGGAAACATGGAATATGAAGAGAAACACAGCAGAG 1560
|||
Db 1501 AATTCCTTCTGCTCCCTTGGAGCTGGAAACATGGAATATGAAGAGAAACACAGCAGAG 1560
|||
Qy 1561 AATTTGTTGATGAAGTTTAAACATTTGCCAAAGACCATGTAAACACAGTTAACTGTA 1620
|||
Db 1561 AATTTGTTGATGAAGTTTAAACATTTGCCAAAGACCATGTAAACACAGTTAACTGTA 1620
|||
Qy 1621 AATTTGTTGATCTTCCACACAGATTTGGAGATATATAGGCTTTCACTTTGAAATGGA 1680
|||
Db 1621 AATTTGTTGATCTTCCACACAGATTTGGAGATATATAGGCTTTCACTTTGAAATGGA 1680
|||
Qy 1681 AAGAGTCCAGATGCTGAGTTTGAACATTTACAGTGTCCCGCAGTCAACACAGAGAGAG 1740
|||
Db 1681 AAGAGTCCAGATGCTGAGTTTGAACATTTACAGTGTCCCGCAGTCAACACAGAGAGAG 1740
|||
Qy 1741 AAAAGAGAAATGGGCTTGAAGCTAGATCTCTGCCATCATCTGATGGGATTTCAACGTG 1800
|||
Db 1741 AAAAGAGAAATGGGCTTGAAGCTAGATCTCTGCCATCATCTGATGGGATTTCAACGTG 1800
|||
Qy 1801 GAAAGATATATGAGGCCACAGCATGATTCACAAAGATCTGAGTCCAGAACACACAC 1860
|||
Db 1801 GAAAGATATATGAGGCCACAGCATGATTCACAAAGATCTGAGTCCAGAACACACAC 1860
|||
Qy 1861 ATCATTGACAATATATCATTTCTGTACTTGGAGAGAAAGAACATGACATTTTGTCTAG 1920
|||
Db 1861 ATCATTGACAATATATCATTTCTGTACTTGGAGAGAAAGAACATGACATTTTGTCTAG 1920
|||
Qy 1921 CTTTCAGAAAACTTCAAGTGTCTCCATCACAGAAATTTATCAGCCACAGAGAGAGTTA 1980
|||
Db 1921 CTTTCAGAAAACTTCAAGTGTCTCCATCACAGAAATTTATCAGCCACAGAGAGAGTTA 1980
|||
Qy 1981 GAGATTGAAGGAGCCCGGGCTGACCTCATTTGAGAGTGTATGAACTTTGAAGATATGCTT 2040
|||
Db 1981 GAGATTGAAGGAGCCCGGGCTGACCTCATTTGAGAGTGTATGAACTTTGAAGATATGCTT 2040
|||
Qy 2041 TGTAAAGTACAGAGAGAAATGGCAAGAAAGAGAGAGAGGCTTTGGGCTGTAGGA 2100
|||
Db 2041 TGTAAAGTACAGAGAGAAATGGCAAGAAAGAGAGAGAGGCTTTGGGCTGTAGGA 2100
|||
Qy 2101 CAGTGTACTATTCAAGACAAAAAACCCAAAGAGAAATGAAAGAAATATCATATTTCTG 2160
|||
Db 2101 CAGTGTACTATTCAAGACAAAAAACCCAAAGAGAAATGAAAGAAATATCATATTTCTG 2160
|||
Qy 2161 AATATGCTGTGCTCCCACTCAAGAGCTTATGATCAAAAGAAACAGTTTGAAGAAATGT 2220
|||
Db 2161 AATATGCTGTGCTCCCACTCAAGAGCTTATGATCAAAAGAAACAGTTTGAAGAAATGT 2220
|||
Qy 2221 GGTTCAGAGTTCTAAAGTGGAGAGATAGCAATGAGTCCCTTATGGCTGCTTCAA 2280
|||
Db 2221 GGTTCAGAGTTCTAAAGTGGAGAGATAGCAATGAGTCCCTTATGGCTGCTTCAA 2280
|||
Qy 2281 AGAAAGAGAAAAATGATGAGAAAAACGTGACAGGCAACCTGTGAGCCATAGCGCTTT 2340
|||
Db 2281 AGAAAGAGAAAAATGATGAGAAAAACGTGACAGGCAACCTGTGAGCCATAGCGCTTT 2340
|||
Qy 2341 CAGCAAGTCCCATACAGTTCTGCAATGTGATGTCAGAGTTGGCTTTCAAGAAATGTAC 2400
|||
Db 2341 CAGCAAGTCCCATACAGTTCTGCAATGTGATGTCAGAGTTGGCTTTCAAGAAATGTAC 2400
|||
Qy 2401 TCGACACCTTGGATCCAAATACGAGGCTGGCATATACCTTCCAAAGAACCTCAAAAC 2460
|||
Db 2401 TCGACACCTTGGATCCAAATACGAGGCTGGCATATACCTTCCAAAGAACCTCAAAAC 2460
|||
Qy 2461 CTGGCAGAGAGGCCAAGAAATCTCTGCTGCAGATTAAGCTGATGTGTTGAAGGCT 2520
|||

```

```

Db 2461 CTGGCAGAGAGGCCAAGAAATCTCTGCTGCAGATTAAGCTGATATGTGTTGAGGCT 2520
Qy 2521 GAAGTCTCACAGGGCTTCTTGCCGAGGACATCCGTTAAATATGTTCGCCACACAG 2580
|||
Db 2521 GAAGTCTCACAGGGCTTCTTGCCGAGGACATCCGTTAAATATGTTCGCCACACAG 2580
|||
Qy 2581 AGTCTGAGCATATAGATGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAC 2640
|||
Db 2581 AGTCTGAGCATATAGATGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAC 2640
|||
Qy 2641 TTTGTTATTTTATAGTGACATGACAGGCTATACCTCACTATTTGTGACATGACACAGAA 2700
|||
Db 2641 TTTGTTATTTTATAGTGACATGACAGGCTATACCTCACTATTTGTGACATGACACAGAA 2700
|||
Qy 2701 TATGTACATCCACAAATATACATCAGAGACCAATGAGACCCCTTGGACAGATCCTTGG 2760
|||
Db 2701 TATGTACATCCACAAATATACATCAGAGACCAATGAGACCCCTTGGACAGATCCTTGG 2760
|||
Qy 2761 AAGGAGATTCGCAAGTGGACGCCCTGTGATTAATCTCTACATCATTTTAAACAGTGTAT 2820
|||
Db 2761 AAGGAGATTCGCAAGTGGACGCCCTGTGATTAATCTCTACATCATTTTAAACAGTGTAT 2820
|||
Qy 2821 GGCTTACCTTGGGTGAACTAACCAAAATATGACATGATGGCTCAAGAGTGGCTTGA 2880
|||
Db 2821 GGCTTACCTTGGGTGAACTAACCAAAATATGACATGATGGCTCAAGAGTGGCTTGA 2880
|||
Qy 2881 ATATATCCCATGGGTTATCTGTATGACATGACTGGGTTATTGAAGACTATGCCACATAC 2940
|||
Db 2881 ATATATCCCATGGGTTATCTGTATGACATGACTGGGTTATTGAAGACTATGCCACATAC 2940
|||
Qy 2941 TAGCATCTTAGTGCCCTTATCTGTCTTATGTCCTTGGGTTGGGATAGATATACAA 3000
|||
Db 2941 TAGCATCTTAGTGCCCTTATCTGTCTTATGTCCTTGGGTTGGGATAGATATACAA 3000
|||
Qy 3001 TGAACACCTTTCAGAGACCTTCTCTCTGTCGAGATGTTCTTAAATCTCCTTACTAGAG 3060
|||
Db 3001 TGAACACCTTTCAGAGACCTTCTCTCTGTCGAGATGTTCTTAAATCTCCTTACTAGAG 3060
|||
Qy 3061 GAGATTAATATTTTGCATATATGAAGAAATTTTCTACTATATATACGAGGCTTTAT 3120
|||
Db 3061 GAGATTAATATTTTGCATATATGAAGAAATTTTCTACTATATATACGAGGCTTTAT 3120
|||
Qy 3121 TTTCTAAATATGATAGATATATAAATGTTAGATTAACGAATGATTTAGATTTCCAG 3180
|||
Db 3121 TTTCTAAATATGATAGATATATAAATGTTAGATTAACGAATGATTTAGATTTCCAG 3180
|||
Qy 3181 AGAATATTTAAAGTCTTATAGATGAATAATATCATCTTGTCTGATTTAAAAAAA 3240
|||
Db 3181 AGAATATTTAAAGTCTTATAGATGAATAATATCATCTTGTCTGATTTAAAAAAA 3240
|||
Qy 3241 AAA 3243
|||
Db 3241 AAA 3243
|||

```

RESULT 2  
ABA93022  
ID ABA93022 standard. cDNA; 3016 BP.

10-APR-2002 (first entry)

Human Interferon-induced tetraspan (IT) protein encoding cDNA.

Human: Interferon-induced tetraspan; IT; cytosolic; anti-HIV;  
immunoprecipitative; antiinflammatory gene therapy; autoimmune disease;  
inflammatory disease; cardiovascular disease; metabolic disease;  
cancer growth inhibition; metastasis; gene; ss.

Homo sapiens.

Key Location/Qualifiers

[illegible]

OY	451	AAAAATAATGAGCTCAGCTGTGTGAAGTCTCTCCAGAAATAGTTTGGCTGTATCTCTACC	510
OY	451	AAAAATAATGAGCTCAGCTGTGTGAAGTCTCTCCAGAAATAGTTTGGCTGTATCTCTACC	510
Db	240	AAAAATAATGAGCTCAGCTGTGTGAAGTCTCTCCAGAAATAGTTTGGCTGTATCTCTACC	299
OY	511	CTGGCTCTCCAGTTTCAGGGAAGCACAAGCAATCTCTGCAAGTGTGCAGAAAATGCTG	570
Db	300	CTGGCTCTCCAGTTTCAGGGAAGCACAAGCAATCTCTGCAAGTGTGCAGAAAATGCTG	359
OY	571	ACTCCTAGAGATAGATTATCATGTCTGGAAAAGATGACCTCCACACATGCTGTGTAGCT	630
Db	360	ACTCCTAGAGATAGATTATCATGTCTGGAAAAGATGACCTCCACACATGCTGTGTAGCT	419
OY	631	GTGGTAATGACCACTTAAGAATCTTCTGCATGGGGAGGCTTGCCCTGGCCTGCTA	690
Db	420	GTGGTAATGACCACTTAAGAATCTTCTGCATGGGGAGGCTTGCCCTGGCCTGCTA	479
OY	691	AAAGCTGTGATTTTGAATAATCCAAAGAAAGAGCAAACTTTTGTGCCAGATATGTATAA	750
Db	480	AAAGCTGTGATTTTGAATAATCCAAAGAAAGAGCAAACTTTTGTGCCAGATATGTATAA	539
OY	751	GTGTACAGTGGTGTAGATGTGTACAGGGAGCGAGGAGGCTTCCCTGCAAACAGATCATC	810
Db	540	GTGTACAGTGGTGTAGATGTGTACAGGGAGCGAGGAGGCTTCCCTGCAAACAGATCATC	599
OY	811	CATGCTGTTGGCCCTCGGTGGATGGAATGGGATTAACAGGAGATGACTGGAAAGCTGCAG	870
Db	600	CATGCTGTTGGCCCTCGGTGGATGGAATGGGATTAACAGGAGATGACTGGAAAGCTGCAG	659
OY	871	AGGGCCATTGTAGATATTTCTGAATATATGTATCTATATAAAATATCTACATTAAAGCAGTA	930
Db	660	AGGGCCATTGTAGATATTTCTGAATATATGTATCTATATAAAATATCTACATTAAAGCAGTA	719
OY	931	GCAATTCACGCTTGAGACTCTGGGAAATTTTTCAGTTCCCTGTAATTTTGTGTCAAGAAGCT	990
Db	720	GCAATTCACGCTTGAGACTCTGGGAAATTTTTCAGTTCCCTGTAATTTTGTGTCAAGAAGCT	779
OY	991	ATTGTAGAGACTATCCGGGTTAGTTTTCGAAGGGAAGCCCAATGATGATTAATTTGAAGAA	1050
Db	780	ATTGTAGAGACTATCCGGGTTAGTTTTCGAAGGGAAGCCCAATGATGATTAATTTGAAGAA	839
OY	1051	ATTCACTGGTGTAGCAATGTAGAGACCTACTGTGTCTTAAAGCTGCTTCAGATATTC	1110
Db	840	ATTCACTGGTGTAGCAATGTAGAGACCTACTGTGTCTTAAAGCTGCTTCAGATATTC	899
OY	1111	ATTCGAGGGAAGATGTGAGTGGGACAAAGAAACCAACCCCTCTTCAATGCAATGGTGTG	1170
Db	900	ATTCGAGGGAAGATGTGAGTGGGACAAAGAAACCAACCCCTCTTCAATGCAATGGTGTG	959
OY	1171	AACAACCTGACCCCTCCAGATTGTGCCAGGGGCACATTGAATGGCAGACGGCAGATGTAAAT	1230
Db	960	AACAACCTGACCCCTCCAGATTGTGCCAGGGGCACATTGAATGGCAGACGGCAGATGTAAAT	1019
OY	1231	GTTAATTTCTGTAANCCACATGATATTACAGTTGACCTGTGTGGCAAAAGTCAATTCTACAA	1290
Db	1020	GTTAATTTCTGTAANCCACATGATATTACAGTTGACCTGTGTGGCAAAAGTCAATTCTACAA	1079
OY	1291	CAACGAGAGTGTGAATGAATGGAATTTCTTGCCACAAGGATTAACAGTTTCAACGG	1350
Db	1080	CAACGAGAGTGTGAATGAATGGAATTTCTTGCCACAAGGATTAACAGTTTCAACGG	1139
OY	1351	TGCCAGTTGTGACGTGTGCACAAAAGATTTTAATCTGTGTGTAATATATATACCATGTA	1410
Db	1140	TGCCAGTTGTGACGTGTGCACAAAAGATTTTAATCTGTGTGTAATATATATACCATGTA	1199
OY	1411	CTGTGGCATTTGAGATTTTCTTAACCTCAGATATTAAACATGCAATGAGAGAGTGTG	1470
Db	1200	CTGTGGCATTTGAGATTTTCTTAACCTCAGATATTAAACATGCAATGAGAGAGTGTG	1259
OY	1471	GAATAATGTACTGTAGCAAAATATATACTTCATTTCTTCTTGCCCTTGGAGCTGGAAAC	1530
Db	1260	GAATAATGTACTGTAGCAAAATATATACTTCATTTCTTCTTGCCCTTGGAGCTGGAAAC	1319

QY 1531 ATGGAATTAAGAGAGAAACAGCAGAGATTGTTGATGAAGTTTAACTTTGGC 1590  
 DB 1320 ATGGAATTAAGAGAGAAACAGCAGAGATTGTTGATGAAGTTTAACTTTGGC 1379  
 QY 1591 AAAGACATGTAAACACCAGTTAACTGTAAATTTTGATCTTTCCACAGATTTGGAG 1650  
 DB 1380 AAAGACATGTAAACACCAGTTAACTGTAAATTTTGATCTTTCCACAGATTTGGAG 1439  
 QY 1651 AATATTAAGCTTTCAGTCTGAATTAAGAGAGTCCAAATCTGTGATTTGAACAT 1710  
 DB 1440 AATATTAAGCTTTCAGTCTGAATTAAGAGAGTCCAAATCTGTGATTTGAACAT 1499  
 QY 1711 TACAGTGTCCCCAGTCAACAGAGAGAGAAAGAAATGGGCTTGAAGTATCT 1770  
 DB 1500 TACAGTGTCCCCAGTCAACAGAGAGAGAGAAAGAAATGGGCTTGAAGTATCT 1559  
 QY 1771 CCTGCATCAATCTGATGGGATTCACCTGGAAGAGATGATAGGCCCCAGCATGATC 1830  
 DB 1560 CCTGCATCAATCTGATGGGATTCACCTGGAAGAGATGATAGGCCCCAGCATGATC 1619  
 QY 1831 CAAGAATCTGAGTCTCCAGAACCCACATCATTTGAATTAATCATATCTGTACCTT 1890  
 DB 1620 CAAGAATCTGAGTCTCCAGAACCCACATCATTTGAATTAATCATATCTGTACCTT 1679  
 QY 1891 GGGAGAAAGAGACATGACATTTTGTCTCAGCTTCAGAAACTTCAAGTCTCCATCACA 1950  
 DB 1680 GGGAGAAAGAGACATGACATTTTGTCTCAGCTTCAGAAACTTCAAGTCTCCATCACA 1739  
 QY 1951 GAAATTAATCAGCCAGAGAGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 2010  
 DB 1740 GAAATTAATCAGCCAGAGAGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1799  
 QY 2011 GAGGTGTTATGACATTAAGATTAAGTCTTTAAAGTCAAGAGAGAGAGAGAGAGAGAG 2070  
 DB 1800 GAGGTGTTATGACATTAAGATTAAGTCTTTAAAGTCAAGAGAGAGAGAGAGAGAGAG 1859  
 QY 2071 AAGGAGCGAGGCTTGTGGCTCTGTAGAGACAGTGAATTCAGACAAACAAACCCAA 2130  
 DB 1860 AAGGAGCGAGGCTTGTGGCTCTGTAGAGACAGTGAATTCAGACAAACAAACCCAA 1919  
 QY 2131 GACGAATTAAGAAATTAATTCATATTTCTGAATGTCCTGTGCTCAACCTCAAGGCTT 2190  
 DB 1920 GACGAATTAAGAAATTAATTCATATTTCTGAATGTCCTGTGCTCAACCTCAAGGCTT 1979  
 QY 2191 CTAGATCAAAAGAAACAGTTGAAAAATGTTGTCAGAGTCTAAAGGTGGAGAGATA 2250  
 DB 1980 CTAGATCAAAAGAAACAGTTGAAAAATGTTGTCAGAGTCTAAAGGTGGAGAGATA 2039  
 QY 2251 GACAATGAGTCTTATGCTCTTTCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2310  
 DB 2040 GACAATGAGTCTTATGCTCTTTCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2099  
 QY 2311 CACAGGCAACCTGTGAGCCTATGCTTTCAGCAAGTCCCATPACCATGTTGCAATGTG 2370  
 DB 2100 CACAGGCAACCTGTGAGCCTATGCTTTCAGCAAGTCCCATPACCATGTTGCAATGTG 2159  
 QY 2371 GATGTCAGAGTGGCTTTCAAAGAAATGTCAGACACCTTGGAGTCCAAATTAAGGAGGT 2430  
 DB 2160 GATGTCAGAGTGGCTTTCAAAGAAATGTCAGACACCTTGGAGTCCAAATTAAGGAGGT 2219  
 QY 2431 GGCATATACCTTCACAGAAACCTCAAAACCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 2490  
 DB 2220 GGCATATACCTTCACAGAAACCTCAAAACCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 2279  
 QY 2491 GCAGATTAAGCTGATGATGTTGTTGAGGCTGAAGTACTACAGAGCTTCTTTCAGGGA 2550  
 DB 2280 GCAGATTAAGCTGATGATGTTGTTGAGGCTGAAGTACTACAGAGCTTCTTTCAGGGA 2339  
 QY 2551 CATCGTTAAATATGTTTCCACACACAGAGCTGAGAGCTATAGATGATGATGACAT 2610  
 DB 2340 CATCGTTAAATATGTTTCCACACACAGAGCTGAGAGCTATAGATGATGATGACAT 2399  
 QY 2611 GTGGTTGACAAATGTCTCAGCCCTGAACCTTTGTTATTTTATGTCATGACAGCTATA 2670

DB 2400 GTGGTTGACAAATGTCTCCAGCCCTGAACCTTGTATTTTATGTCATGAGCAGCTATA 2459  
 QY 2671 CCTCAGTATTTTGTGACATGACACCAGAGATATGTACAGTCCAAAGATTTACTATCAGAGA 2730  
 DB 2460 CCTCAGTATTTTGTGACATGACACCAGAGATATGTACAGTCCAAAGATTTACTATCAGAGA 2519  
 QY 2731 CCAATGAGACCTTGTGACAGATCTTGGAGGGATTCGCAAGTGGAGGCCCTGTAT 2790  
 DB 2520 CCAATGAGACCTTGTGACAGATCTTGGAGGGATTCGCAAGTGGAGGCCCTGTAT 2579  
 QY 2791 TATCTTACATCATTTTAAACAGTGTATGCTTACCTTGGGTGAACTAAACAAATTA 2850  
 DB 2580 TATCTTACATCATTTTAAACAGTGTATGCTTACCTTGGGTGAACTAAACAAATTA 2639  
 QY 2851 TGACCATGAGTGTCCAAAGAGTGTGATATATCCCATGAGTTATCTGTATGAGCTG 2910  
 DB 2640 TGACCATGAGTGTCCAAAGAGTGTGATATATCCCATGAGTTATCTGTATGAGCTG 2699  
 QY 2911 ACTGGTTATTAAGAGACTACCCACATACATGATCTTATGCTTATCTTATCTTAT 2970  
 DB 2700 ACTGGTTATTAAGAGACTACCCACATACATGATCTTATGCTTATCTTATCTTAT 2759  
 QY 2971 GTCTTGGGGTTGGGTAGTAGATACCAATGAACACATTCAGAGACCTTCTCTCTT 3030  
 DB 2760 GTCTTGGGGTTGGGTAGTAGATACCAATGAACACATTCAGAGACCTTCTCTCTT 2819  
 QY 3031 GCAGTGTCTTTTAAATCTCTTACTAGAGAGATTAATTTTGCATATATTAAGAGAA 3090  
 DB 2820 GCAGTGTCTTTTAAATCTCTTACTAGAGAGATTAATTTTGCATATATTAAGAGAA 2879  
 QY 3091 TTTTCTAGTATATTAAGCAGGCTTTTATTTCTAAATGATGATTAATTAATGTT 3150  
 DB 2880 TTTTCTAGTATATTAAGCAGGCTTTTATTTCTAAATGATGATTAATTAATGTT 2939  
 QY 3151 AGGATTAACGAATGATTTAGATTTTCCAGAGATTAATTAAGTCTTATGATGA 3210  
 DB 2940 AGGATTAACGAATGATTTAGATTTTCCAGAGATTAATTAAGTCTTATGATGA 2999  
 QY 3211 ATAAATCATCTTGTCT 3227  
 DB 3000 ATAAATCATCTTGTCT 3016

RESULT 3  
 ID ABA83077 standard; DNA; 2598 BP.  
 XX  
 AC ABA83077;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Human transcription factor TRFX-104 coding sequence.  
 XX  
 DE Human; transcription factor; TRFX; cell proliferative disease;  
 KW autoimmune disease; inflammation; neurological disease;  
 KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;  
 KW neuroprotective; antiinflammatory; gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200172777-AZ.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 13-MAR-2001; 2001WO-US08117.  
 XX  
 PR 13-MAR-2000; 2000US-0188986.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PI Hillman JL, Baughn MR, Yue H, Lai P, Lu DM, Patterson C;  
 PI Azimzal Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;





```
Db 61 GCTGTGATAGCTCTCACGGGAGCAGGAGGCTTCCTCGCAACAGATCATCTGCT 120
Qy 817 GTTGGGCTGTGGTGGATGGAATAAACAGGAGTGTACTGGAAAGCTGCAGAGGGCC 876
Db 121 GTTGGGCTGTGGTGGATGGAATAAACAGGAGTGTACTGGAAAGCTGCAGAGGGCC 180
Qy 877 ATTGTAAGTATTCGATTAATGTCATCTATATAAAATCTACATTAAGACAGTACAT 936
Db 181 ATTGTAAGTATTCGATTAATGTCATCTATATAAAATCTACATTAAGACAGTACAT 240
Qy 937 CCAGCCTTGAGCTCTGGGATTTTTCAGTTCCTCTGTAATTTGGTACAAAGACTATTG 996
Db 241 CCAGCCTTGAGCTCTGGGATTTTTCAGTTCCTCTGTAATTTGGTACAAAGACTATTG 300
Qy 997 GAGACTATCCGGGGTTAGTTTGCAAGGGAGCAATGATGATTAATTTGAAGAATTCAC 1056
Db 301 GAGACTATCCGGGGTTAGTTTGCAAGGGAGCAATGATGATTAATTTGAAGAATTCAC 360
Qy 1057 CTGGTGAAGCAATGAGACCCCTACTGTTGCTGCTTTAAAGCTGCTTCAAGATTCCT 1116
Db 361 CTGGTGAAGCAATGAGACCCCTACTGTTGCTGCTTTAAAGCTGCTTCAAGATTCCT 420
Qy 1117 GGGAAAGTGAAGCTGGGACAAAGAACCCCTTCTTTCATGCAATGTCGTGAACAC 1176
Db 421 GGGAAAGTGAAGCTGGGACAAAGAACCCCTTCTTTCATGCAATGTCGTGAACAC 480
Qy 1177 CTGACCTTCAGATTTGTCAGGGCCACATTTGAATGGCAGCGCAGATGTAATTTGTA 1236
Db 481 CTGACCTTCAGATTTGTCAGGGCCACATTTGAATGGCAGCGCAGATGTAATTTGTA 540
Qy 1237 TCTGTAAACCCACATGATTAATTTACAGTTTGACCTGTGSCAAAGTCAATTTCAACA 1296
Db 541 TCTGTAAACCCACATGATTAATTTACAGTTTGACCTGTGSCAAAGTCAATTTCAACA 600
-Qy 1297 GAGGTGAATGAATGGGAATTTCTTCCACAAAGGCTAAACGTTTCAACGCTCCAG 1356
Db 601 GAGGTGAATGAATGGGAATTTCTTCCACAAAGGCTAAACGTTTCAACGCTCCAG 659
Qy 1357 TTGGTACTGTGCACAAAGATTTAACTTGTCTGTAATATATATATACATGTACTGTG 1416
Db 660 TTGGTACTGTGCACAAAGATTTAACTTGTCTGTAATATATATATACATGTACTGTG 719
Qy 1417 CATTCAAGATTTCTTAAACCTCAGATATTAACATGCAATGAAGAGTGTTGGAAAA 1476
Db 720 CATTCAAGATTTCTTAAACCTCAGATATTAACATGCAATGAAGAGTGTTGGAAAA 779
Qy 1477 TGCATTGAGAAAAATTAATCTCATCTTCTTCTGCTCCCTGGGAGTGCAGAAATGGA 1536
Db 780 TGCATTGAGAAAAATTAATCTCATCTTCTTCTGCTCCCTGGGAGTGCAGAAATGGA 839
Qy 1537 ATAAAGAGAAACAGCAGCAGAGATTTTGTGATGAAGTTTAACTTGGCCAAAGAC 1596
Db 840 ATAAAGAGAAACAGCAGCAGAGATTTTGTGATGAAGTTTAACTTGGCCAAAGAC 899
Qy 1597 CATGTAAACACAGTTAACTGTAAATTTGTGATCTTTCCACACAGATTTGGAGATAT 1656
Db 900 CATGTAAACACAGTTAACTGTAAATTTGTGATCTTTCCACACAGATTTGGAGATAT 959
Qy 1657 AAGGCTTTCAGTTCTGAATAGGCAAGAGTCCAAAGATGCGAGTTGAGCAATTAAGT 1716
Db 960 AAGGCTTTCAGTTCTGAATAGGCAAGAGTCCAAAGATGCGAGTTGAGCAATTAAGT 1019
Qy 1717 GTCCCCAGTCAACAGAGAGAGAAAAAAGAAATGGCTTGAAGCTAGATCTCTGCC 1776
Db 1020 GTCCCCAGTCAACAGAGAGAGAAAAAAGAAATGGGCTTGAAGCTAGATCTCTGCC 1079
Qy 1777 ATCAATCTGATGGGATTCACAGTGAAGAGATGTAAGGCCCAAGCATGATCAAGA 1836
Db 1080 ATCAATCTGATGGGATTCACAGTGAAGAGATGTAAGGCCCAAGCATGATCAAGA 1139
Qy 1837 ATCTGTAGTCTCCAGAACCAACACATATGAGATAATATATCTGTACTTTGGGAGA 1896
Db 1140 ATCTGTAGTCTCCAGAACCAACACATATGAGATAATATATCTGTACTTTGGGAGA 1199
Qy 1897 AAGAACATGACATTTTGTCTCAGCTTCAGAAAACTCAAGTGTCTCATCAGAAAT 1956
Db 1200 AAGAACATGACATTTTGTCTCAGCTTCAGAAAACTCAAGTGTCTCATCAGAAAT 1259
Qy 1957 ATCAGCCAGAGAGAGCAGAGTATGAGATGAAGAGACCCGGGCTGACCTATTGAGGTG 2016
Db 1260 ATCAGCCAGAGAGAGCAGAGTATGAGATGAAGAGACCCGGGCTGACCTATTGAGGTG 1319
Qy 2017 GTTATGACATTTGAAGTATGCTTTTGTAAAGTACAGAGGAATTCGCAAGAAAAAGAG 2076
Db 1320 GTTATGACATTTGAAGTATGCTTTTGTAAAGTACAGAGGAATTCGCAAGAAAAAGAG 1379
Qy 2077 CGAGGCTTTGGGCTGTTAGACAGTGAACATTTAGCAACAAAAACCCAAAGACAA 2136
Db 1380 CGAGGCTTTGGGCTGTTAGACAGTGAACATTTAGCAACAAAAACCCAAAGACAA 1439
Qy 2137 ATGAAGAAAAATATCATATTTTGTAAATGCTGTGCTTCAACTCAAGAGCTTCTAGAT 2196
Db 1440 ATGAAGAAAAATATCATATTTTGTAAATGCTGTGCTTCAACTCAAGAGCTTCTAGAT 1499
Qy 2197 CAAAAGAACAGTTTGAATAATGTTGCTGAGTTCTAAAGTGAGAGATAGACAT 2256
Db 1500 CAAAAGAACAGTTTGAATAATGTTGCTGAGTTCTAAAGTGAGAGATAGACAT 1559
Qy 2257 GAGGTCTTATGCTGCTGCTTCAAGAAAGAAAGAAATGATGAGAAAGAACTGCACAG 2316
Db 1560 GAGGTCTTATGCTGCTGCTTCAAGAAAGAAAGAAATGATGAGAAAGAACTGCACAG 1619
Qy 2317 CAACCTGTGAGCCATGAGCTGTTTTCAGCAGTCCCATACAGTTTCTCAATGTGTAATGC 2376
Db 1620 CAACCTGTGAGCCATGAGCTGTTTTCAGCAGTCCCATACAGTTTCTCAATGTGTAATGC 1679
Qy 2377 AGAGTTGGCTTCAAAAGATGTAAGTGTGACACTTTGGCATCAAAATGCGAGGTGGCATA 2436
Db 1680 AGAGTTGGCTTCAAAAGATGTAAGTGTGACACTTTGGCATCAAAATGCGAGGTGGCATA 1739
Qy 2437 TACTTACCAAGAACCTCAGAAAACTTGGCAGAGAGAGCCAAAGAAATCTGTGTCAGAT 2496
Db 1740 TACTTACCAAGAACCTCAGAAAACTTGGCAGAGAGAGCCAAAGAAATCTGTGTCAGAT 1799
Qy 2497 AAGCTGATCTATGTTGTGAGGCTGAAGTACTCAGAGGCTTCTTCCACAGGACATCCG 2556
Db 1800 AAGCTGATCTATGTTGTGAGGCTGAAGTACTCAGAGGCTTCTTCCACAGGACATCCG 1859
Qy 2557 TTAAATATGTTGCCCCACAGCTGAGTCCGAGGCTATAGATGCTCATGACAGTGGTT 2616
Db 1860 TTAAATATGTTGCCCCACAGCTGAGTCCGAGGCTATAGATGCTCATGACAGTGGTT 1919
Qy 2617 GACATGCTCCAGACCCCTGAACCTTTGTTATTTTATGTTGAGTGCAGGCTATACCTCAG 2676
Db 1920 GACATGCTCCAGACCCCTGAACCTTTGTTATTTTATGTTGAGTGCAGGCTATACCTCAG 1979
Qy 2677 TATTTTGGACATGACACCCAGAAATATGTACAGTACACAAATTAATCATGAGCAATG 2736
Db 1980 TATTTTGGACATGACACCCAGAAATATGTACAGTACACAAATTAATCATGAGCAATG 2039
Qy 2737 AGACCCCTTGCACAGATCCCTTGGAGGGAGTTGSCAAGTGGCAGGCCCTGTTGATTAATCT 2796
Db 2040 AGACCCCTTGCACAGATCCCTTGGAGGGAGTTGSCAAGTGGCAGGCCCTGTTGATTAATCT 2099
Qy 2797 CTACATCATTTTAAACAGCTGTATGCTTACCTTGGGTAACCTAACCAATTAATGACCA 2856
Db 2100 CTACATCATTTTAAACAGCTGTATGCTTACCTTGGGTAACCTAACCAATTAATGACCA 2159
Qy 2857 TCGATGCTCAAGAGTGGCTTGAATATATCCCATGAGGCTTATCTGTATGACATGACCTGGG 2916
Db 2160 TCGATGCTCAAGAGTGGCTTGAATATATCCCATGAGGCTTATCTGTATGACATGACCTGGG 2219
Qy 2917 TTATTTGAAGAGCTAGCAGCATATCTAGCATCTAGTCCCTTATATCTTATATGCTTGTG 2976
Db 2220 TTATTTGAAGAGCTAGCAGCATATCTAGCATCTAGTCCCTTATATCTTATATGCTTGTG 2279
```

OY	2977	GGTGGGGTAGAGAGACTACCAAAATGAACACCTTCAGACCTCCTCCTTGCACTT	3036
Db	2280	GGGTGGGGTAGAGAGATACCAAAATGAACACCTTCAGACCTCCTCCTTGCACTT	2339
OY	3037	GTCCTTAACTCCTCTTACTAGAGAGATAAATATTTTGCAATATGAGAAATTTTTC	3096
Db	2340	GTCCTTAACTCCTCTTACTAGAGAGATAAATATTTTGCAATATGAGAAATTTTTC	2399
OY	3097	TAGATATATACGCAGGCCCTTTATTTTCTTAAATGATGATAGTATAAATGTTAGATA	3156
Db	2400	TAGATATATACGCAGGCCCTTTATTTTCTTAAATGATGATAGTATAAATGTTAGATA	2459
OY	3157	ACAGAAATGATTTAGATTTTCCAGAGAAATATATAAGGCTTTAGGTAGAAATAAT	3216
Db	2460	ACAGAAATGATTTAGATTTTCCAGAGAAATATTTAAGGCTTTAGGTAGAAATAAT	2519
OY	3217	CATCTTTGCTGATTTAAAAAATAAAAA	3243
Db	2520	CATCTTTGCTGATTTAAAAAATAAAAA	2546

RESULT 5	
AAK52513/c	
ID AAK52513	standard; cDNA; 1695 BP.

AC AAK52513;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 2042.

**KW** Human; cytokine; cell proliferation; cell differentiation; gene therapy;

**KW** tissue growth factor; immunomodulatory; cancer; leukaemia;

XX

XX

XX

XX

XX

PR 27-APR-2000; 2000US-0560875.

PR 19-JUL-2000; 2000US-0620325.

PR 15-SEP-2000; 2000US-0663561.

PR 30-NOV-2000; 2000US-0728422.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT

PI Xue AJ, Yang Y, Wejhrman T

DR WPI; 2001-476283/51.

XX

PT useful in diagnosis and gene

PS Claim 1; Page 4437; 6221pp;

CC The invention relates to pol

cytokine, cell proliferation

polynucleotides and polypeptides

	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC	(AAH80020) are omitted as the relevant pages from the sequence listing
CC	were missing at the time of publication.
XX	
XX	Sequence 1695 BP; 452 A; 379 C; 323 G; 541 T; 0 other;
Query Match	47.1%; Score 1527.8; DB 22; Length 1695;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1529; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Dy	1713 CAGTGTCCCCCGCTGACACCAGAGAGAGAAAAAGAAAATGGGCTTTGAAGTCATGCC 1772
Dd	1531 CAAGAAGCCCCCGTAGCACCCAGAGAGAGAAAAAGAAAATGGGCTTTGAAGTCATGCC 1472
Oy	1773 TGCCATCATATCGAATGGGATTCACAGTGGAAGAGATGTATGAGGCCCAAGCATGATCCA 1832
Dd	1471 TGCCATCATATCGAATGGGATTCACAGTGGAAGAGATGTATGAGGCCCAAGCATGATCCA 1412
Oy	1833 AAGAAATCCTGAGTCTCCAGAACCCACATCATATTGAGAAATAATCATATTTCTGTACTTTGG 1892
Dd	1411 AAGAAATCCTGAGTCTCCAGAACCCACATCATATTGAGAAATAATCATATTTCTGTACTTTGG 1352
Oy	1893 GAGAAAGGAACAATGCATATTTGTCTCAGCTTCAGAAAACCTCAAGTGTCTCATACAGCA 1952
Dd	1351 GAGAAAGGAACAATGCATATTTGTCTCAGCTTCAGAAAACCTCAAGTGTCTCATACAGCA 1292
Oy	1953 AATTAATCAGCCCGAGAGAGACAGAGATTAGAGATTAAAGAGCCCGGCTGCATCTATGA 2012
Dd	1291 AATTAATCAGCCCGAGAGAGACAGAGATTAGAGATTAAAGAGCCCGGCTGCATCTATGA 1232
Oy	2013 GGTTGTTATGAACATTTGAAGATATGCTTTGTAAGTACAGAGAGAAATGGCAAGSAAAA 2072
Dd	1231 GGTTGTTATGAACATTTGAAGATATGCTTTGTAAGTACAGAGAGAAATGGCAAGSAAAA 1172
Oy	2073 GAGAGGAGGCTTTTGGCGCTGCTTAGAGACAGTGGACTATTACAGCAACAAAAAACCAGCA 2132
Dd	1171 GAGAGGAGGCTTTTGGCGCTGCTTAGAGACAGTGGACTATTACAGCAACAAAAAACCAGCA 1112
Oy	2133 CGAAATGGAAGAAATATCATATTTCTGAAATGTCTGCTGCCCTCCAACTCAAGACTTCT 2192
Dd	1111 CGAAATGGAAGAAATATCATATTTCTGAAATGTCTGCTGCCCTCCAACTCAAGACTTCT 1052
Oy	2193 AGATCAAAAAGAAACAGTTTTGAAAAATGTGGTTGCAAGCTTCTAAAGGTTGGAAGATAGA 2252
Dd	1051 AGATCAAAAAGAAACAGTTTTGAAAAATGTGGTTGCAAGCTTCTAAAGGTTGGAAGATAGA 992
Oy	2253 CAATGAGTCTTTATGTGCTGCTTTCAAAGAAAGAAATGATGGAAGAAAACTGCA 2312
Dd	991 CAATGAGTCTTTATGTGCTGCTTTCAAAGAAAGAAATGATGGAAGAAAACTGCA 932
Oy	2313 CAGGCAACCTGTGAGGCAATGAGGCTTTTCAGCAAGTCCCATACGATCTGCAATGTGGT 2372
Dd	931 CAGGCAACCTGTGAGGCAATGAGGCTTTTCAGCAAGTCCCATACGATCTGCAATGTGGT 872
Oy	2373 ATGCAAGATGTGCTTTCAAAGAAATGTACTCGACACCTTGCGATTCCAAAAATAGAGAGCTGG 2432
Dd	871 ATGCAAGATGTGCTTTCAAAGAAATGTACTCGACACCTTGCGATTCCAAAAATAGAGAGCTGG 812
Oy	2433 CATATATCTTACCAAGAAACCTCAAAAACCTGGCGAGAGAAAGCCAAAGAAATCTTGCTGCG 2492
Dd	811 CATATATCTTACCAAGAAACCTCAAAAACCTGGCGAGAGAAAGCCAAAGAAATCTTGCTGCGCG 752
Oy	2493 AGATAAGCTGATCTATGTGCTTTGAGGCTGAATACTCAAGGCTTCTTCTGCGAGGAGACA 2552
Dd	751 AGATAAGCTGATCTATGTGCTTTGAGGCTGAATACTCAAGGCTTCTTCTGCGAGGAGACA 692
Oy	2553 TCCTGTTAAATATTTGTCCCACCAAGTGCCTGAGACTATAGATGGTTCATGACAGTGT 2612

Dd	651	TCCGTTAAATATATGTGTCCCCACCACTAGTCCTGGAGCTATACATGTCATACACTGT	632
Qy	2613	GATTGCACAATGTCTTCCAGCACCCTGAACCCCTTGTTATTTTTAGTGGCATGCAAGGCTATACC	2672
Dd	631	GGTTTCAACATGCTCTCCACGCCCTGAAACCCCTTGTTATTTTTAGTGGCATGCAAGGCTATACC	572
Qy	2673	TCAGTATTTTGGGACATGCACCACAGGAATATGTACAGTCCACAGAATTAATCATAGAAC	2732
Dd	571	TCAGTATTTTGGGACATGCACCACAGGAATATGTACAGTCCACAGAATTAATCATAGAAC	512
Qy	2733	AATGAGACCCCTTTGGACAGCATCCCTGGAGGGGATATGCGAAGTGGACGCCCTGTTGATTA	2792
Dd	511	AATGAGACCCCTTTGGACAGCATCTCTGGAGGGGATATGCGAAGTGGACGCCCTGTTGATTA	452
Qy	2793	ATCTCTACATCATYTTTAAACAGCTGGTATGGCCTTACCTTGGGTGAACTAACCAATAATG	2852
Dd	451	ATCTCTACATCATYTTTAAACAGCTGGTATGGCCTTACCTTGGGTGAACTAACCAATAATG	392
Qy	2853	ACCATCATGGCTCAAAGATGGCTTGAAATATATCCCATGGGTTATCTGATGSACTGAC	2912
Dd	391	ACCATCATGGCTCAAAGATGGCTTGAAATATATCCCATGGGTTATCTGATGSACTGAC	332
Qy	2913	TGGGTTATGGAAGSACTAGGACAAATACTAGACATCTTAGTGCCCTTATPMTGCTCTTATGT	2972
Dd	331	TGGGTTATGGAAGSACTAGGACAAATACTAGACATCTTAGTGCCCTTATPMTGCTCTTATGT	272
Qy	2973	CTTGGGGTTGGGGTAGTAGATACCAATGGAACACTTTCAGSACCTCCCTCCTCTTGC	3032
Dd	271	CTTGGGGTTGGGGTAGTAGATACCAATGGAACACTTTCAGSACCTCCCTCCTCTTGC	212
Qy	3033	AGTTCTCTTTAATCTCCTTTACTAGAGAGATAAATATTTTGCATATATGAAAGAAATT	3092
Dd	211	AGTTCTCTTTAATCTCCTTTACTAGAGAGATAAATATTTTGCATATATGAAAGAAATT	152
- Qy	3093	TTTCTAGATATPAACGCGAGCCCTTTATTTCTAAAAATGATGATAGATATAAAATGTTAG	3152
Dd	151	TTTCTAGATATPAACGCGAGCCCTTTATTTCTAAAAATGATGATAGATATAAAATGTTAG	92
Qy	3153	GATAACAGAAATGATTTTATGATTTTCCAGAGATATTTAATAGTCTTTAGGTATGAAT	3212
Dd	91	GATAACAGAAATGATTTTATGATTTTCCAGAGATATTTAATAGTCTTTAGGTATGAAT	32
Qy	3213	AAATCATCTTTGTCTGATTTAAAAAAAAA 3243	
Dd	31	AAATCATCTTTGTCTGATTTAAAAAAAAA 1	
<b>RESULT 6</b>			
AKS1529	ID	AAKS1529 standard; cDNA; 1747 BP.	
XX	AAKS1529;		
AC	06-NOV-2001	(first entry)	
DT	Human polynucleotide seq ID NO 74.		
XX			
DE			
XX			
KW	Human: cytokine; cell proliferation; cell differentiation; gene therapy		
KM	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		
KW	tissue growth factor; immunomodulatory; cancer; leukemia;		
KM	nervous system disorder; arthritis; inflammation; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200157190-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	05-FEB-2001; 2001WO-US04098.		
XX			
PR	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
PR	20-JUN-2000; 2000US-0598075.		

PR	19-JUL-2000;	2000US-0620325.
PR	01-SEP-2000;	2000US-0634336.
PR	15-SEP-2000;	2000US-0633561.
PR	20-OCT-2000;	2000US-0693325.
PR	30-NOV-2000;	2000US-0728422.
XX	(HYSE-) HYSEQ INC.	
PA	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y	
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;	
DR	WPI: 2001-476283/51.	
XX	P-PSDB: AAM78396.	
PT	Nucleic acids encoding polypeptides with cytokine-like activities,	
PT	useful in diagnosis and gene therapy -	
PS	Claim 1; Page 675-677; 6221pp; English.	
XX	The invention relates to polynucleotides (AAK51456-AAK53435) and the	
CC	encoded polypeptides (AAM78323-AAKM0302) that exhibit actively elating	
CC	cytokine, cell proliferation or cell differentiation or which may indu	
CC	production of other cytokines in other cell populations. The	
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines	
CC	peptide therapy. The polypeptides have various cytokine-like activities	
CC	e.g. stem cell growth factor activity, hematopoiesis regulating	
CC	activity, tissue growth factor activity, immunomodulatory activity and/or	
CC	activin/inhibin activity and may be useful in the diagnosis and/or	
CC	treatment of cancer, leukemia, nervous system disorders, arthritis an	
CC	inflammation.	
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666	
CC	(AAM80020) are omitted as the relevant pages from the sequence listing	
CC	were missing at the time of publication.	
XX	Sequence 1747 BP; 546 A; 342 C; 399 G; 460 T; 0 other;	
SO		
Query Match	47.1%; Score 1526.2; DB 22; Length 1747;	
Best Local Similarity	99.8%; Pred. No. 0;	
Matches 1526:	Conservative 0; Mismatches 3; Indels 0; Gaps	
OY	1713 CAGTGTCCCCCAGTCAACGAGAGAGAAGAAAATGGCGCTTGAACTAGATCTCC	177Z
Db	217 CAAGTCCCCCAGTCACACGAGAGAGGAAGAAAATGGCGCTTGAACTAGATCTCC	276
OY	1773 TGCCATCATCTGTATGGGATTTCACACGTGGAGAAGATGTATGAGGCCCCACGCATGGATCCA	1832Z
Db	277 TGCCATCATCTGTATGGGATTTCACACGTGGAGAAGATGTATGAGGCCCCACGCATGGATCCA	336
OY	1833 AAGAACCTCGATGTCACAGAACACCACACATCATTTAGATTAATCATATTCTGTACTCTGG	1892Z
Db	337 AAGAACTCTGAGTCTCCAGAACCCACCATCATTTAGATTAATCATATTCTGTACTCTGG	396
OY	1893 GAGAAAGAACATGACATTTTTGTCTCAGCTTCGAAAAACTCAAGTGTCTCCATCACAGA	1952Z
Db	397 GAGAAAGAACATGACATTTTTGTCTCAGCTTCGAAAAACTCAAGTGTCTCCATCACAGA	456
OY	1953 AATTATCAGCCCACAGAGAGACAGATTTAGAGATTTGAAGAGCCCGGCTGACCTATTGA	2012Z
Db	457 AATTATCAGCCCACAGAGAGACAGATTTAGAGATTTGAAGAGCCCGGCTGACCTATTGA	516
OY	2013 GGTTGTTATGAACATTTAATAATATGCTTTTAAGTACAGAGAGAAATGGCAAGGAAAAA	2072Z
Db	517 GGTTGTTATGAACATTTAATAATATGCTTTTAAGTACAGAGAGAAATGGCAAGGAAAAA	576
OY	2073 GGAGCGAGGCGCTTTGGCGCTGTTTAGACAGTGACATATTGACACAACAAAACCCAGA	2132Z
Db	577 GGAGCGAGGCGCTTTGGCGCTGTTTAGACAGTGACATATTGACACAACAAAACCCAGA	636
OY	2133 CGAATTAAGAAATAATATCATATTCTGAAATGTCCTGTGGCTCCACATCAAGACCTCT	2192Z
Db	637 CGAATTAAGAAATAATATCATATTCTGAAATGTCCTGTGGCTCCACATCAAGACCTCT	696

QY 2193 AGATCAAAAGAAACAGTTTGAAAAATGTGTTGAGCGTTCTCTAAAGTGAGAGAGATAGA 2252  
 |||||  
 Db 697 AGATCAAAAGAAACAGTTTGAAAAATGTGTTGAGCGTTCTCTAAAGTGAGAGAGATAGA 756  
 |||||  
 QY 2253 CAATGAGGTCCTTATGCGTCCCTTCAAAGAAAGAAAAATGATGAGAAAAAAGCTGCA 2312  
 |||||  
 Db 757 CAATGAGGTCCTTATGCGTCCCTTCAAAGAAAGAAAAATGATGAGAAAAAAGCTGCA 816  
 |||||  
 QY 2313 CAGCAACCTGTGAGCCATAGGCTGTTCAGCAAGTCCCATACAGTTCTGCAATGTGT 2372  
 |||||  
 Db 817 CAGCAACCTGTGAGCCATAGGCTGTTCAGCAAGTCCCATACAGTTCTGCAATGTGT 876  
 |||||  
 QY 2373 ATGCAGAGTTGGCTTTCAAAGAAATGTACGACACCTGGCCGATCCAAATACGGAGCTGG 2432  
 |||||  
 Db 877 ATGCAGAGTTGGCTTTCAAAGAAATGTACGACACCTGGCCGATCCAAATACGGAGCTGG 936  
 |||||  
 QY 2433 CATATACTTCACCAAGAACTCAAAAACCTGGCAGAGAGCCCAAGAAAATCTCTGTCG 2492  
 |||||  
 Db 937 CATATACTTCACCAAGAACTCAAAAACCTGGCAGAGAGCCCAAGAAAATCTCTGTCG 996  
 |||||  
 QY 2493 AGATGAGCTGATCTATGTGTGTTGAGGCTGAAAGTACTCAGAGCTTCTTCTGCGAGGACA 2552  
 |||||  
 Db 997 AGATGAGCTGATCTATGTGTGTTGAGGCTGAAAGTACTCAGAGCTTCTTCTGCGAGGACA 1056  
 |||||  
 QY 2553 TCCGTTAAATATGTGTTCCCGCCACACCTGAGTCCCTGAGCTATAGATGGTCATGACAGTGT 2612  
 |||||  
 Db 1057 TCCGTTAAATATGTGTTCCCGCCACACCTGAGTCCCTGAGCTATAGATGGTCATGACAGTGT 1116  
 |||||  
 QY 2613 GGTGACAAATGTCTCCAGCCCTGAAACCTTTGTTATTTTATAGTGGCATGACAGCTATAC 2672  
 |||||  
 Db 1117 GGTGACAAATGTCTCCAGCCCTGAAACCTTTGTTATTTTATAGTGGCATGACAGCTATAC 1176  
 |||||  
 QY 2673 TCAGTATTTGGGACATGACATGACCCAGAGATATGTACAGTCACANAGTTACTCATCGAGAC 2732  
 |||||  
 Db 1177 TCAGTATTTGGGACATGACATGACCCAGAGATATGTACAGTCACANAGTTACTCATCGAGAC 1236  
 |||||  
 QY 2733 AATGAGACCTTTTGACAGACATCCCTTGAGAGGATTTCCAAATGGCAGCCCTGTGATTA 2792  
 |||||  
 Db 1237 AATGAGACCTTTTGACAGACATCCCTTGAGAGGATTTCCAAATGGCAGCCCTGTGATTA 1296  
 |||||  
 QY 2793 ATCTCTACATCATTTTAAACAGCTGTATGCGCTTACCTTGGGTGAATCAACCAAAATATG 2852  
 |||||  
 Db 1297 ATCTCTACATCATTTTAAACAGCTGTATGCGCTTACCTTGGGTGAATCAACCAAAATATG 1356  
 |||||  
 QY 2853 ACCATCGATGGCTCAAAAGAGTGGCTTGAAATATATCCCATGGGTTATCTGTATGAGCTGAC 2912  
 |||||  
 Db 1357 ACCATCGATGGCTCAAAAGAGTGGCTTGAAATATATCCCATGGGTTATCTGTATGAGCTGAC 1416  
 |||||  
 QY 2913 TGGGTATTTGAAAGAGTACGACATACATAGCATCTTATGTCCTTATCTGTATATGT 2972  
 |||||  
 Db 1417 TGGGTATTTGAAAGAGTACGACATACATAGCATCTTATGTCCTTATCTGTATATGT 1476  
 |||||  
 QY 2973 CTGCGGTTGGGGTATGATATACCAAAATGAACACTTTCAGAGACCTTCTCTTGGC 3032  
 |||||  
 Db 1477 CTGCGGTTGGGGTATGATATACCAAAATGAACACTTTCAGAGACCTTCTCTTGGC 1536  
 |||||  
 QY 3033 AGTGTCTTTAAATCTCTTAACTAGAGAGATTAATATTTTGCATATATAGAAAAAT 3092  
 |||||  
 Db 1537 AGTGTCTTTAAATCTCTTAACTAGAGAGATTAATATTTTGCATATATAGAAAAAT 1596  
 |||||  
 QY 3093 TTTCTAGTATTAACGAGCGCTTTATTTCTAAATAGATGATGTATTAAGATGTAG 3152  
 |||||  
 Db 1597 TTTCTAGTATTAACGAGCGCTTTATTTCTAAATAGATGATGTATTAAGATGTAG 1656  
 |||||  
 QY 3153 GATAACAGAAATGATTTAGATTTTCCAGAGAAATATTAAGAGCTTATAGTATGATGAT 3212  
 |||||  
 Db 1657 GATAACAGAAATGATTTAGATTTTCCAGAGAAATATTAAGAGCTTATAGTATGATGAT 1716  
 |||||  
 QY 3213 AAATCATCTTGTGATTAAGAAAAA 3243  
 |||||  
 Db 1717 AAATCATCTTGTGATTAAGAAAAA 1747  
 |||||

RESULT 7  
 AAD00630  
 ID AAD00630 standard; cDNA: 3024 BP.  
 XX  
 AC AAD00630;  
 XX  
 DT 29-AUG-2000 (first entry)  
 XX  
 DE Mouse B-aggressive lymphoma (BAL) protein-short form cDNA.  
 KW Mouse; B-aggressive lymphoma; BAL; tumour; malignancy;  
 KW differential expression; DLB-CL; Diffuse large B-cell lymphoma;  
 KW cytosolic; vaccine; treatment; diagnosis; non-Hodgkin's lymphoma;  
 KW cellular adhesion; sarcoma; carcinoma; myeloma; ss.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FH 5'UTR 1..170  
 FH CDS /\*tag= a  
 FT 171..2651  
 FT /\*tag= b  
 FT /product= "BAL protein-short form"  
 FT 171..2648  
 FT /\*tag= c  
 FT /note= "This region is specifically claimed in Claim 1  
 as SEQ ID NO:6"  
 FT 3'UTR 2652..3024  
 FT /\*tag= c  
 FT misc\_feature 2295..2739  
 FT /\*tag= d  
 FT /note= "This region is 99% identical to Soares 2NBWF  
 Mus musculus cDNA clone 1446050"  
 FT  
 PN W0200026231-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PE 29-OCT-1999; 99WO-US25439.  
 XX  
 PR 29-OCT-1998; 98US-0106383.  
 PR 30-OCT-1998; 98US-0106448.  
 XX  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 XX  
 PI Shipp M, Aguilar R, Yakushijin Y;  
 XX  
 DR WPI: 2000-365563/31.  
 DR P-PSDB: AAY71083.  
 XX  
 PT New B-aggressive lymphoma nucleic acid for identifying cells exhibiting  
 PT or predisposed to malignancies such as lymphoma, sarcoma, carcinoma and  
 PT myeloma -  
 PS  
 PS  
 XX  
 Claim 1; Fig 2; 151pp; English.  
 CC The present sequence is a cDNA encoding short form of B-aggressive  
 CC lymphoma (BAL) protein obtained from mouse (Balb-c) spleen by 5' and 3'  
 CC RACE using a 418 bp clone as an anchor from Soares mouse mammary gland  
 CC having homology to human sequence and primers from mouse EST database.  
 CC BAL was identified to be differentially expressed in DLB-CL (Diffuse  
 CC large B-cell lymphoma), the most common non-Hodgkin's lymphoma, by  
 CC differential display technique. It was found to be more abundant in  
 CC tumours from patients with high risk fatal DLB-CL disease than low risk  
 CC cured disease. BAL transcripts are most abundant in lymphoid organs e.g.  
 CC spleen, foetal liver and peripheral blood, and several non-haematopoietic  
 CC organs e.g. heart and skeletal muscle. BAL is involved in modulation of  
 CC cellular adhesion and aggressiveness/severity of malignancy such as  
 CC DLB-CL. BAL nucleic acid and protein are useful for identifying cells  
 CC exhibiting or predisposed to malignancies such as lymphoma, sarcoma,  
 CC carcinoma and myeloma. BAL proteins are useful as immunogens to raise  
 CC anti-BAL antibodies. BAL modulators or BAL molecules can be used for  
 CC prophylactic and therapeutic treatment of a subject susceptible to or

CC having a disorder associated with aberrant BAL expression or activity,  
CC such as non-Hodgkin's lymphoma. It is also useful in vaccine preparation.  
XX

Sequence 3024 BP; 902 A; 635 C; 725 G; 762 T; 0 other;  
Query Match 41.3%; Score 1340; DB 21; Length 3024;  
Best Local Similarity 70.5%; Pred. No. 0;  
Matches 2053; Conservative 0; Mismatches 735; Indels 122; Gaps 15;

367 CCTACAGAGCTCAGAGAGCTGCTCTGAGAAAAACATATAGTGGCAATTCCTCAT 426  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 192 CCGCGCGAAGAGACCGACCAATCTCTGTAAGAACATATAGATGCAATTCCTCAT 251  
427 AACCAATGACTTCAAAATTTTAAATAATATAGCGCTAGCTGTGTGAATCTCCAG 486  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 252 AAACACAATGCTTCGAAATTTTAAAGAGCAATGAGAGTACGATAGTGAAGCTCCCAA 311  
487 AATAAGTTGGCTGTATCTACCTACCTGCTCCAGTTCCAGTGAAGAGCAACAGC---AA 543  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 312 AATAAGTTGGATGATCATCTACCTGAGCTGTACCACTAGACAGGAGACGCTCTCT 371  
544 TCTGTCAAGTGTTCAGAAAAATGCTGACCTAGATAGTATATCAGTTCGAAAGAT 603  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 372 GCTCAGAGAGTCTTCAGAAAGAGACCTGATCCCTGGGATAGATTATCTGTCTGGAAGAT 431  
604 GACCTACACACATGCTGTGATGCTGTGTGTAATGCAACCAATGAAGATCTTGCAT 663  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 432 GACCTTACACAGACAGCTGTGTGATGCTGTGTGTAACGACCAATGAAGCTTTGCAAT 491  
664 GGGGAGAGCTGGCCCTGGCCCTGTGTAAAGCTGTGTGATTTGAAATCCAAAGAGAGC 723  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 492 GGAAGTGGCTGGCCGCGAAGCTGTGTAAAGCTGTGTGTAATCCAAAGAGAGC 551  
724 AAACAGTTTGTTCAGATATGTAAGTGTGACCTGTGTGATAGTGTACAGGGAGCA 783  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 552 AAAGAGATCTTGTCCAAAGCTGTGTAATATCTGATTTGTGATTCGCTATCACCGGTGC 611  
784 GGGAGGCTCCCTGCAAAACATCATCATGCTGTGGGCTGGGTGATGTAAGGAT 843  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 612 GGGAGACTCTCTGCTGCAATTTGATTTATCCATGCGGTGGAGCCTCGGTGACAGTTACAGAC 671  
844 AAACAGGATGTACTGGAAGCTGACAGAGGCGCATTTGATGTTTGTGATTTATGTCATC 903  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 672 AGCCAGACACTATCGAATTCGAAATTTGCCATTAGGAACATTTGATTTATGTCACC 731  
904 TATAAAAATCTCAGATTAAGACAGTACGATTCAGACCTTGAAGCTGTGGATTTTTCAG 963  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 732 AATATATGATCTACGATTTAAGACAGTACGATTTCCAGCCTGAGCTCTGGAATTTTCAG 791  
964 TTCCCTCTGAATTTGTGTACAAAAGACTATTTGTAGAGACTATCCGGGTTTACTTTGCAAGG 1023  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 792 TTCCCTCTGAATTTGTGTACAAAAGACTATTTGTAGAACTATTCGGGCTTTATTTCCAGAC 851  
1024 AAACCAATGATGATATTTGAAAGAAATTCACCTGTGTGACAAATGAGACCTACTGTT 1083  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 852 AAACCAATGATGATATTTGAAAGAAATTCATCTGTGTGCAATGAGGACCCCAAGCTT 911  
1084 GTGCGCTTTAAAGCTGTGTGCAATTCATCTGAGGGAAGAG-----TGAGCTGTGGA 1134  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 912 GGTCTCTTAAATCCGCTGTGAGAAAGCATCTTGTGAGGAGGAGACCTGAGCTTTTGGGGGGGT 971  
1135 CAAGAAACCAACCCCTCTT-----TCATGCAATGCTGTGTAACAACCTGACCTCCAG 1188  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 972 CCAAGAAACGACCTGCTGTCCACCATGACATCTTCGCAATGCGGCGGGGCTGACCTCCAG 1031  
1189 ATTGTCCAGGCGCACATTTGATGAGAGGCGAGATGTAATTTGTAATTTCTGTAACCCA 1248  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1032 ATTGTCCAGGCGTGTATTTGAAATGCAAAACAGATGTAATTTGTAATTTCTGATCATG 1091  
1249 CATGATATTTAGTGTGACCTGTGGAAGAGTCAATCTTACAAAGAGAGAGTGAAGT 1308  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1092 CAGGATTTTAAATCAGAGAGTGTGACAGTCAATCTTATGACAAAGAGGGGTTGAAGT 1151

1309 AATTCGATTTCTTGCCCAAGAGCTAAACAGATTTCAACGGTCCAGTTGTAATGCTC 1368  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1152 GAAAGGAACTT-----GACAAGGTTAACTCTCCACAGATTTATCAAGAGGTGGGCTC 1205  
1369 ACAAAAGATTTTAACTGTTCTGTAATATATATACCATGTACTGTCGATTCAGATTT 1428  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1206 ACAAAAGATTTTAAATTTGCTCTGATATGCTATCTTCCATGTGTCATGATGCTCCAAATC 1265  
1429 CTTAACTTCAGATTTTAAATAATGCAATGCAATGCAAGAGTGTGTAAGAAATGCAATGAGCA 1488  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1266 AACAAATACAGATATTTGAAAGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1325  
1489 AATATATCTTCCATTTCTTCTCTGCTGCTGGGACGACGTAAGCAATGTAAGAAAGAA 1548  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1326 GATATATATTTCAATTTCTTCTGCTGCTGGAAGAGATGATGATTTGAAAGAGAGT 1385  
1549 ACAGCAGAGAGATTTGTTGATGAGTTTAAATTTGCAATTTGCCAAAGACCTGTAAGAACAC 1608  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1386 ACAGCAGCTCAGATATTTGTTGAGAGATTTTGCATTTCTTAAAGACCAAGAGAAAA 1445  
1609 CAGTTAAGCTTAAATTTTGTGATCTTCCAAAGATTTGAGATATTAAGCTTTCAAT 1668  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1446 AGCTTAAGCTTAAAGATTTGATCTTTCCAGTATGATGAGAGCTGTAAGAGATTTTAT 1505  
1669 TCTGAATGSCAAGAGGCTCAAGATGCTGAGTTTGAACATTTACAGT-----GTC 1719  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1506 GCTGAATATGCAAAAGAGTCCAAAGAGCTGAATCTCAGCGTATATGATGCTTTAGCC 1565  
1720 CCCCAGTCAACAGAGAGGAAAGAAATGGGCTTGAACCTGATGATCTCTGCATC 1779  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1566 CTGCAAGTGTCCAGTGGGAGGCAAGAGAGGCGGCTTGAACGATGATCTCTGCATC 1825  
1780 AATCTGATGATTTCAACGCTGAGAGAGATATGAGGCGCCACCATGATCCAAAGATC 1839  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1626 AATCTGATGATTTAAAGTGGAGAGATGCTGAGGCGCCAGGAATGATTTAAAGGTTG 1885  
1840 CTGAGTCTCCAGAACCAACATCATTTGAGATATATTTGTAATCTTGGGAGAGAG 1899  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1686 CTGCTCTCCCTGAGCACCAACATCATTTGAGATATATCTATCTATCTGTTGAAAAA 1745  
1900 GAACATGACATTTTGTCTGCTGCAAGAAACCTTCAAGTGTCTCCATCAGAAATTAATC 1959  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1746 GAGCAGAGAGCTGCTGTGCTGCTCAGACCAACAGCAAGAGATCTCATTTCCAGAGCTGTC 1805  
1960 AGCCAGAGAGAGCAGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1806 AGTCCAAAG 1865  
2020 ATGACATTTGAGATATGCTTTTAAAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1866 ATGAGATTTGAGATATGCTTTTAAAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1925  
2080 GGCCTTTGGGCGCTGTAGAGAGAGTGTGATTTTCAAGCAAAAGAGAGAGAGAGAGAG 2139  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1926 AATCTTTGAGAGCTTTGAG 1982  
2140 AAAGAAATATCATTTTGAATGTCTGCTCTCAACATCAAGAGCTTTTGAATCA 2199  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 1983 GAGAGATGTCACATTTTCAAGATATCCAGCATCATTTAATCTAGAGACTTCAGAGACCA 2042  
2200 AAGAAACAGTTTGAAGATGCTTTGACAGTTCTAAAGGTGAGAGAGATGACAAATGAG 2259  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 2043 AAGAAACAGTTTGAAGATGCTTTGACAGTTCTAAAGGTGAGAGAGATGACAAATGAG 2102  
2260 GTCTCTATGCTCTCTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2319  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 2103 GTCTCTATGCTCTCTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2162  
2320 CTGTGAGCATATGAGCTGTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2379  
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
DB 2163 TCTGGAG 2222  
2380 GTTGGCTTTCAAGAAATGTAATCTGACAGCTTTGCAATCAAAATATGAGAGCTGACATATAC 2439

```

Db 2223 GTGGCTTCCACAGATGATGATCGACATCTATACCCAGTTTATGAGCCGGCATATAT 2282
Qy 2440 TTCACCAAGAACCTCAAAAACCTGGCAGAGAGCCAAAGAAAATCTGCTGCAGATAG 2499
Db 2283 TTCACCAAGAACCTCAAAAACCTGACAGACAAGGTCAAGAAAACCTCAAGCACAGACAG 2342
Qy 2500 CTGATCTGATGTTGAGGCTGAGTACTCAACAGGCTCTCTGCGCAGGACATCCGTTA 2559
Db 2343 CTATCTCTGTTGTTGAGGCGAGAACTACTCAGAGGCTCTCTGCGGATATTCCTCA 2402
Qy 2560 AATATTTGTCGCCACCACTGAGTCCCTGAGCTATAGATGTCATGACAGTGTGTTGAC 2619
Db 2403 AATATCATCCCTCCACCACTGAGTCCCTGAGGCTTATAGATGTCATGACAGGCTATTTGAC 2462
Qy 2620 AATGCTCCAGCCCTGAAACCTTTGTTATTTTATGCTGACAGGCTATACGTAT 2679
Db 2463 AATGTTCCAGCCCTGAAACCACTGTTGTTTATGCTGACAGGCTATGCCCTGTGAC 2522
Qy 2680 TTGTCACATGACCCAGC-----AATATGTCAGTCAACAGATTAC 2721
Db 2523 TTGTGACATTTGCACACAGATAGACATTCACAGCATCCGATGTGTCACAGACTAC 2582
Qy 2722 TCATCAGAGCAATGAGACCCCTTTCACAGACATCCCTGAGGGATTCGCAAGTGCAGC 2781
Db 2583 TCATCAGAGCAAGATGCTCTCTCGCTGCGTGCAGTCCGGAATGGGTCTTAATGAGCAGC 2642
Qy 2782 CCTGTGATTAATCTCTCATCATTTTAAAGAGCTGTATGCGCTTACCTGGGTAACTA 2841
Db 2643 TCTGT---TTAGTGTCTCATCATGTTTAAACAAGAGG-----GTTCGAGAACCTG 2693
Qy 2842 ACCAATATATGACCATTCATGAGCTCAAGAGAGTGGCTGAATATATCCATGGTATATCG 2901
Db 2694 ACAAATGATATAAT-----ACAGGTTAACTG 2720
Qy 2902 TATGACACTGAGTGGTTATTTGAAGAGCTAGCCACATACATCAATAGTCCCTTATC 2961
Db 2721 TTCAGAAATGATGGGTCTACTAAAGCAGCAGCACACATAGCATATAGTCCCTTGTGC 2780
Qy 2962 TGTCTTATGCTTGGGTTGGGGTAGTAGATACCAATGAACACTTTCAGACCTTC 3021
Db 2781 TTTACCTCT-----GGGCTGTGACGAGGAGATGCCAGTAAACCTTCTCAGCTGCTTT 2834
Qy 3022 CTTCCTCTGACATGTTCTTAAATCCTTACTAGAGAGATATAATTTTGGATATA 3081
Db 2835 TCT-ATTGACATCTTTATCTCTTCTCTATAGGTGACAGCAAGAACTTTATATAGA 2893
Qy 3082 ATGAAGAAATTTTCTAGTATATAAGCAGGCTTTTATTTTCTAAATGATGATAGTAT 3141
Db 2894 ACAAGGATATTTTTT-----CAAGCTGTATTTTCTTAAATGATAGACACAA 2942
Qy 3142 AAAATGTTAGATTAACAGATGATTTTAGATTTCCAGAGAAATATATTAAGTCTTAA 3201
Db 2943 C-----TAGACAAACAGGATTTTCAAGTTTCTATATAT-TTATTAAGTCTTTGG 2994
Qy 3202 GGTATGAAATTAATCATCTTGTCTGATT 3231
Db 2995 GATATCCAATTAATCACCCTTGTCTGACT 3024

```

```

RESULT 8
ABV23030
ID ABV23030 standard; cDNA: 854 BP.
XX
AC ABV23030;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 23021.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX

```

```

OS Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX
XX 16-MAR-2000; 2000US-189862P.
XX
XX 25-MAY-2000; 2000US-207454P.
XX
XX 09-JUN-2000; 2000US-211314P.
XX
XX 18-JUL-2000; 2000US-219007P.
XX
XX 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI, 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer
XX
XX Claim 1; Page 4103-4104; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for:
XX (a) assessing whether a patient is afflicted with prostate cancer;
XX (b) monitoring the progression of prostate cancer in a patient;
XX (c) assessing the efficacy of a test compound to inhibit prostate
XX cancer in a patient;
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound;
XX (g) determining whether prostate cancer has metastasized in a patient;
XX (h) assessing the aggressiveness or indolence of prostate cancer in a
XX patient;
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 854 BP; 272 A; 171 C; 183 G; 227 T; 1 other:
XX
XX Query Match 24.4%; Score 792; DB 23; Length 854;
XX Best Local Similarity 98.2%; Pred. No. 3.6e-205;
XX Matches 823; Conservative 0; Mismatches 10; Indels 5; Gaps 2;
Qy 898 GTCATCTATAAAATACATACATTAAGACAGTAGCAATTCAGCCTTGAGCTCTGGATT 957
Db 20 GTCATCTATAAAATACATACATTAAGACAGTAGCAATTCAGCCTTGAGCTCTGG--- 76
Qy 958 TTTCAGTTCCTCTGAAATTTGTGTACAAGACTATTTAGAGACTATCCGGGTTAGTTTG 1017
Db 77 -CAGCGGTGGTCTGTAATTTGTGTACAAGACTATTTGTAGAGACTATCCGGGTTAGTTTG 135
Qy 1018 CAAGGGAAGCAATGATGATTAATTTGAAAGAAATTCACCTGCTGAGCAATAGAGACCT 1077
Db 136 CAAGGGAAGCAATGATGATTAATTTGAAAGAAATTCACCTGCTGAGCAATAGAGACCT 195
Qy 1078 ACTGTTGCTGCTTTAAACCTGCTTACAGAAATTCATCTGAGGAAGTGAAGTGGAGCAA 1137
Db 196 ACTGTTGCTGCTTTAAACCTGCTTACAGAAATTCATCTGAGGAAGTGAAGTGGAGCAA 255
Qy 1138 GAAACACCCCTTCTTTCATGCAATGCTGTGAACAACCTGACCCCTCAGATTGTCCAG 1197
Db 256 GAAACACCCCTTCTTTCATGCAATGCTGTGAACAACCTGACCCCTCAGATTGTCCAG 315
Qy 1198 GGCACATTGAATGGCAGAGGAGAGATGTAATTTGTAATTTCTGTAACCCACATGATATT 1257
Db 316 GGCACATTGAATGGCAGAGGAGAGATGTAATTTGTAATTTCTGTAACCCACATGATATT 375

```

QY 1258 ACAGTTGGACCTGTGGCAAGTCAATTTCTACAACAGCAGAGATTGAATGAATCGGAA 1317  
DB 376 ACAGTTGGACCTGTGGCAAGTCAATTTCTACAACAGCAGAGATTGAATGAATCGGAA 435  
QY 1318 TTCTTGGCCAAAGGCTTAACAGTTTCAACGGTCCAGTTGGTACTGGTCACAAAAGGA 1377  
DB 436 TTCTTGGCCAAAGGCTTAACAGTTTCAACGGTCCAGTTGGTACTGGTCACAAAAGGA 495  
QY 1378 TTTAACCTGTCTGTAAATATATATACATGTACTGTGGCATTGAGAAATTTCTAAACCT 1437  
DB 496 TTTAACCTGTCTGTAAATATATATACATGTACTGTGGCATTGAGAAATTTCTAAACCT 555  
QY 1438 CACATATTTAAACATGATGAAGAGTGTGGAAAAATGCAATTGACAAAATATATACT 1497  
DB 556 CACATATTTAAACATGATGAAGAGTGTGGAAAAATGCAATTGAGCAAAAATATATACT 615  
QY 1498 TCATTTCTCTTCTGCTGCTGGAGTGGAAACATGGAATTAAGAAAGAAAGAGAGA 1557  
DB 616 TCATTTCTCTTCTGCTGCTGGAGTGGAAACATGGAATTAAGAAAGAAAGAGAGA 675  
QY 1558 GAGA-TTTTGTGTGATGAAGTTTAAACATTTGCCAAAGACATGTAAACACCAGTTAAC 1616  
DB 676 GAGATTTTGTGTGATGAAGTTTAAACATTTGCCAAAGACATGTAAACACCAGTTAAC 735  
QY 1617 TGTAAATTTGTGATCTTTTCCAAACAGATTGGAGATATATAGGCTTTCAGTTCTGAAT 1676  
DB 736 TGTAAATTTGTGATCTTTTCCAAACAGATTGGAGATATATAGGCTTTCAGTTCTGAAT 795  
QY 1677 GGCAAAGAGGTCCAAAGATGCTGAGTTTGAACATTTACAGTGTCCCGCAGCAACCGCA 1734  
DB 796 GGCAAAGAGGTCCAAAGATGCTGAGTTTGAACATTTACAGTGTCCCGCAGCAACCGCA 853

RESULT 9

-ABV28866  
ID ABV28866 standard; cDNA: 854 BP.  
XX ABV28866;

XX 16-SEP-2002 (first entry)  
XX  
XX

DE Human prostate expression marker CDNA 28857.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

PR 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX  
PS Claim 1; Page 6082-6083; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX

SO Sequence 854 BP; 272 A; 171 C; 183 G; 227 T; 1 other;

Query Match 24.4%; Score 792; DB 23; Length 854;

Best Local Similarity 98.2%; Pred. No. 3,6e-205;

Matches 823; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

QY 898 GTCATCTATATAAATCTACATTAAGACAGTGAATTCAGCTTGGAGATT 957  
DB 20 GTCATCTATATAAATCTACATTAAGACAGTGAATTCAGCTTGGAGATT 76  
QY 958 TTTCAGTCCCTCTGTAATTTGTGTACAAAGACTATGTAGAGACTATCCGGGTAGTTG 1017  
DB 77 -GACGGGTGGCTGTGAATTTGTGTACAAAGACTATGTAGAGACTATCCGGGTAGTTG 135  
QY 1018 CAAGGAAAGCCAAATGATGATTAATTTGAAAGAAATTCACCTGGTGAAGAGACCT 1077  
DB 136 CAAGGAAAGCCAAATGATGATTAATTTGAAAGAAATTCACCTGGTGAAGAGACCT 195  
QY 1078 ACTGTGCTGCTTTAAACCTGTTAGAAATTCATCTGAGAGAGAGTACTGGAGACA 1137  
DB 196 ACTGTGCTGCTTTAAACCTGTTAGAAATTCATCTGAGAGAGAGTACTGGAGACA 255  
QY 1138 GAACCCACCCCTCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1197  
DB 256 GAACCCACCCCTCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 315  
QY 1198 GGCACATTTGAATGCGAGAGCGAGATGTAATTTGTAATTTCTGTAAACCCAGATGATT 1257  
DB 316 GGCACATTTGAATGCGAGAGCGAGATGTAATTTGTAATTTCTGTAAACCCAGATGATT 375  
QY 1258 ACAGTTGGACCTGTGGCAAGTCAATTTCTACAACAGCAGAGTGAATGAATCGGAA 1317  
DB 376 ACAGTTGGACCTGTGGCAAGTCAATTTCTACAACAGCAGAGTGAATGAATCGGAA 435  
QY 1318 TTCTTGGCCAAAGGCTTAACAGTTTCAACGGTCCAGTTGGTACTGGTCACAAAAGGA 1377  
DB 436 TTCTTGGCCAAAGGCTTAACAGTTTCAACGGTCCAGTTGGTACTGGTCACAAAAGGA 495  
QY 1378 TTTAACCTGTCTGTAAATATATATACATGTACTGTGGCATTGAGAAATTTCTAAACCT 1437  
DB 496 TTTAACCTGTCTGTAAATATATATACATGTACTGTGGCATTGAGAAATTTCTAAACCT 555  
QY 1438 CAGATATTTAAACATGATGAAGAGTGTGGAAAAATGCAATTGAGCAAAAATATATACT 1497  
DB 556 CAGATATTTAAACATGATGAAGAGTGTGGAAAAATGCAATTGAGCAAAAATATATACT 615  
QY 1498 TCATTTCTCTTCTGCTGCTGGAGTGGAAACATGGAATTAAGAAAGAAAGAGAGA 1557  
DB 616 TCATTTCTCTTCTGCTGCTGGAGTGGAAACATGGAATTAAGAAAGAAAGAGAGA 675  
QY 1558 GAGA-TTTTGTGTGATGAAGTTTAAACATTTGCCAAAGACATGTAAACACCAGTTAAC 1616  
DB 676 GAGATTTTGTGTGATGAAGTTTAAACATTTGCCAAAGACATGTAAACACCAGTTAAC 735  
QY 1617 TGTAAATTTGTGATCTTTTCCAAACAGATTGGAGATATATAGGCTTTCAGTTCTGAAT 1676  
DB 736 TGTAAATTTGTGATCTTTTCCAAACAGATTGGAGATATATAGGCTTTCAGTTCTGAAT 795

```
QY 1677 GCCTAAGAGGTCTCAAGATGCTAGTTGACAAATTACAGTGTCCCGCAGTACACAGA 1734
    |||||||
Db 796 GGCCTAAGAGGTCTCAAGATGCTAGTTGACAAATTACAGTGTCCCGCAGTACACAGA 853

RESULT 10
AAH72506
ID AAH72506 standard; cDNA; 822 BP.
XX
AC AAH72506;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 3780.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN W0200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 718; 1051ip; English.
XX
CC The invention relates to novel genes (AAH6727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 822 BP; 228 A; 177 C; 211 G; 198 T; 8 other;

Query Match 19.2%; Score 622.8; DB 22; Length 822;
Best Local Similarity 86.1%; Pred. No. 4.5e-159;
Matches 759; Conservative 0; Mismatches 17; Indels 106; Gaps 2;

QY 121 GCGGGAGCGGGGCGGAGTGGGCGCACCATTATCTGGAACACTATGCTTATGTTGA 180
    |||||
Db 45 GAGGAGACCGGCGCTGCGGAAGTGGGCGCACCATTATCTGGAACACTATGCTTATGTTGA 104

QY 181 AGCGCAAAAGGGAATTAACATTTAAAGACTCCCGGGGACCTGGAGATGACTTTTCC 240
    |||||||
Db 105 AGCGCAAAAGGGAATTAACATTTAAAGACTCCCGGGGACCTGGAGATGACTTTTCC 164

QY 241 ATGTGGCGGAGAGCAGCAGCTTACATTAATAAATCAGGTAGGATTACTCGCTCTCACTC 300
    |||||||
Db 165 ATGTGGCGGAGAGCAGCAGCTTACATTAATAA----- 197

QY 301 TTGTTTCAAGAAAGTCTTGTCTCAGATCTTTCCTCAGTGGAGAAAGGGAATACAGAAAG 360
```

```
Db 198 ----- 197
QY 361 TGTCTCCCTCAACAGTGTCTCAGAGACGTGCTCTGTGGAGAAACTATAGTTGGCAATT 420
    |||||||
Db 198 -----TCAGAGACTGTGCTCTTGTGAGAGAAACTATAGTTGGCAATT 239

QY 421 CCCATTAAACCAATTAAGCTTCAAAATTTTAAATAATGAGCTGCTGCTGTGAAGTC 480
    |||||||
Db 240 CCCATTAAACCAATTAAGCTTCAAAATTTTAAATAATGAGCTGCTGCTGTGAAGTC 299

QY 481 CTCAGATAATGATTTGGCTGTATCTTAACCTGCTGCTCAGTTCAGGAAGCAACAGC 540
    |||||||
Db 300 CTCAGATAATGATTTGGCTGTATCTTAACCTGCTGCTCAGTTCAGGAAGCAACAGC 359

QY 541 AATCTCTGCAAGTGTTCAGAAAATGCTGCTCTGTGAGATAGGATATACGCTGGA 600
    |||||||
Db 360 AATCTCTGCAAGTGTTCAGAAAATGCTGCTCTGTGAGATAGGATATACGCTGGA 419

QY 601 GATGACCTCACACACATGCTGTGATGCTGTGATGAGAGCCAAATGAAGATCTCTG 660
    |||||||
Db 420 GATGACCTCACACACATGCTGTGATGCTGTGATGAGAGCCAAATGAAGATCTCTG 479

QY 661 CATGGGGAGGCTGTGGCCCTGTGGCTGTAAAGCTGTGATTTGAATCCAAGAGAG 720
    |||||||
Db 480 CATGGGGAGGCTGTGGCCCTGTGGCTGTAAAGCTGTGATTTGAATCCAAGAGAG 539

QY 721 AGCAACAGTTTGTGTCAGATATGTAAGTGTACACTGCTGTGATAGCTGTACCGGA 780
    |||||||
Db 540 AGCAACAGTTTGTGTCAGATATGTAAGTGTACACTGCTGTGATAGCTGTACCGGA 599

QY 781 GCAGGGAGGCTTCCCTGCAACACAGATCATCTGTGTGGGCTCGGTGATGAATGG 840
    |||||||
Db 600 GCAGGGAGGCTTCCCTGCAACACAGATCATCTGTGTGGGCTCGGTGATGAATGG 659

QY 841 GATAAACAGGATGTACTGAAAGCTGCAGAGG-CCATTGTAGATTTCTGAATATGT 899
    |||||||
Db 660 GATAAACAGGATGTACTGAAAGCTGCAGAGGCTGATGTAGTATCCGGAATATGT 719

QY 900 CATCTATTAATAATCTACATTAGACAGTATGCAATTCACACCTTGAGCTCTGGGATTT 959
    |||||||
Db 720 CATCTATTAATAATCTACATTAGACAGTATGCAATTCACACCTTGAGCTCTGGGATTT 779

QY 960 TCAGTCCCTCTGAATTTGTGTACAAAGACTATTTGTAGAGAC 1001
    |||||||
Db 780 TCAGTCCCTCTGAATTTGTGTACAAAGACTATTTGTAGAGAC 821

RESULT 11
ABY30165
ID ABY30165 standard; cDNA; 820 BP.
XX
AC ABY30165;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 30156.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
```

PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI: 2001-662795/76.  
DR  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
PS  
PS Claim 1; Page 6531; 11750pp; English.  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 820 BP; 226 A; 178 C; 211 G; 197 T; 8 other;  
Query Match 18.8%; Score 609; DB 23; Length 820;  
Best Local Similarity 85.1%; Pred. No. 2,6e-155;  
Matches 750; Conservative 0; Mismatches 25; Indels 106; Gaps 2;  
QY 121 GCGGGAGCGCGGCGCAGAGTGGGCCACATATCTGGAAATCACTATGCTTTGA 180  
DB 45 GAGGAGAGCGGCGCTCGCGGAGTGGGCCACATATCTGGAAATCACTATGCTTTGA 104  
QY 181 ACCGCAAAAGGGAATTAACATTTAAAGACTCCCGGGGAGCTGGAGATGACTTTTCC 240  
DB 105 ACCGCAAAAGGGAATTAACATTTAAAGACTCCCGGGGAGCTGGAGATGACTTTTCC 164  
QY 241 ATGTGGCCCGGAGCAGACACTTACAATGAAAAATCAGTAGATTACTCGCTCAGTC 300  
DB 165 ATGTGGCCCGGAGCAGACACTTACAATGAAAAATCAGTAGATTACTCGCTCAGTC 197  
QY 301 TTGTTTCAGAAAGTCTTTGCTCAGATCTTCTCAGTGGAGAAAGGGAATACAGAA 360  
DB 198 ----- 197  
QY 361 TGTCTCCCTACAGAGTGTGCTGCTTTGGAAAACTATATGTTGGCAATT 420  
DB 198 -----TCAGAGACTGTGCTCTTGGAAAACTATATGTTGGCAATT 239  
QY 421 CCCATTAAACCAATGACTCTTCAAAATTTTAAAAAATATAGAGCGAGCTGTGAAGTC 480  
DB 240 CCCATTAAACCAATGACTCTTCAAAATTTTAAAAAATATAGAGCGAGCTGTGAAGTC 299  
QY 481 CTCAGAAATAGTTGGCTGTATCTCTACCTCGTCTCCAGTTCCAGAGGCAACAGC 540  
DB 300 CTCAGAAATAGTTGGCTGTATCTCTACCTCGTCTCCAGTTCCAGAGGCAACAGC 359  
QY 541 AAATCTCTCAAGTGTTCGAAAAAATGCTGACTCCTAGATAGAGTTATCATGCTTGGAAA 600  
DB 360 AAATCTCTCAAGTGTTCGAAAAAATGCTGACTCCTAGATAGAGTTATCATGCTTGGAAA 419  
QY 601 GATGAGCTACACACATGCTGTGATGCTGTGGTGAATGAGCCAAATAGCTTTTCG 660  
DB 420 GATGAGCTACACACATGCTGTGATGCTGTGGTGAATGAGCCAAATAGCTTTTCG 479  
QY 661 CATGGGGAGGCGCTGCGCCCTGCTGATAAAGCTGTGATTGAAATCCAGAAAGAG 720

DB 480 CATGGGGAGGCGCTGCGCCCTGCTGATAAAGCTGTGATTGAAATCCAGAAAGAG 539  
QY 721 AGCAAAAGTGTTCGAGATATGTAAGTGTGAGTGTGAGATAGCTTCACGGGA 780  
DB 540 AGCAAAAGTGTTCGAGATATGTAAGTGTGAGTGTGAGATAGCTTCACGGGA 599  
QY 781 GCAGGAGGCTTCCCTGCAAAACAGATCATCATGCTGTTGGGCTCGGTGATGAAATGG 840  
DB 600 GCAGGAGGCTTCCCTGCAAAACAGATCATCATGCTGTTGGGCTCGGTGATGAAATGG 659  
QY 841 GATTAACAGGAGTGTACTGGAAGAGCTGCAGAGGCGCATTTGTAGTATTCGATTATGTC 900  
DB 660 GATTAACAGGAGTGTACTGGAAGAGCTGCAGAGGCGCATTTGTAGTATTCGATTATGTC 718  
QY 901 ATCTATTAATAATACATACATTAAGACAGTAGCAATTCGACCTTGAGCTTGGATTTT 960  
DB 719 ATCTATTAATAATACATACATTAAGACAGTAGCAATTCGACCTTGAGCTTGGATTTT 778  
QY 961 CAGTTCCTCTGCAATTTGTGTACAAAGACTATTTAGAGAC 1001  
DB 779 CAGTTCCTCTGCAATTTGTGTACAAAGACTATTTAGAGAC 819  
RESULT 12  
ID AAH72659  
AAH72659 standard; cDNA; 666 BP.  
XX  
AC AAH72659;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 3933.  
XX  
KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
XX WPI: 2001-375006/39.  
DR  
XX New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
PS  
PS Claim 1; Page 785; 1051pp; English.  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 666 BP; 191 A; 139 C; 183 G; 152 T; 1 other;

	Query Match	15.14%	Score 489.4	DB 22	Length 666	
	Best Local Similarity	84.6%	Pred. No. 9.1e-123			
	Matches 608	Conservative 0	Mismatches 6	Indels 105	Gaps 1	
QY	136 GCACAGAGTGGGCGCCACCATATCTGGAACCTACAGTCTATGCTTTGAAGGCCAAGAGGANT	195				
Db	52 GAAAAAGTGGGCGCCACCATATCTGGAACCTACAGTCTATGCTTTGAAGGCCAAGAGGANT	111				
QY	196 AACATTTTAAAGACTCCCCGGGGGACCTGGAGGATGGACTTTTCCATGGTGGCCGGAGCA	255				
Db	112 AACATTTTAAAGACTCCCCGGGGGACCTGGAGGATGGACTTTTCCATGGTGGCCGGAGCA	171				
QY	256 GCAGCTTTCATGAAAAATTCAGGTAGGATTACCTCGCTCTCACTCTGTTTCAGAAAGTTC	315				
Db	172 GCAGCTTTCATGAAAA-----	189				
QY	316 TTTGCTCAGATCTTCCATCAGTGGAGAAAGGGGANTACGAAGANTGTCCCTACAG	375				
Db	190 -----	189				
QY	376 TGCTCAGAGACTGTCCTCTTGGAGAAAACTATAGTTGGCAATTTCCATTAACCAAT	435				
Db	190 ---TCAGAGACTGTCCTCTTGGAGAAAACTATAGTTGGCAATTTCCATTAACCAAT	246				
QY	436 GACTTCAAAATTTTAAAAAATATAGAGCGTCAGCTGTGTGAAGTCCCTCCAGATAAGTTT	495				
Db	247 GACTTCAAAATTTTAAAAAATATAGAGCGTCAGCTGTGTGAAGTCCCTCCAGATAAGTTT	306				
QY	496 GCGTCTATCTCAACCTGGTCTCTCCAGTTCAGGAAGGCAACAGCAATCTCGCAAGT	555				
Db	307 GCGTCTATCTCAACCTGGTCTCTCCAGTTCAGGAAGGCAACAGCAATCTCGCAAGT	366				
QY	556 TTCAGAAAAAATCTCACTCTCTAGATAGATATCAGTCTGGAAGATGACCTCACACCA	615				
Db	367 TTTACAAAAAATCTCACTCTCTAGATAGATATCAGTCTGGAAGATGACCTCACACCA	426				
QY	616 CATGCTGTTGATGCTGTGGTGAATCAGCCAAATGAAGATCTTCTGCATGGGGAGGCCCTG	675				
Db	427 CATGCTGTTGATGCTGTGGTGAATCAGCCAAATGAAGATCTTCTGCATGGGGAGGCCCTG	486				
QY	676 GCCCTGGCCCTGGTAAAGCTGGTGGATTGAAATCCAAAGAGAGCAACAGTTGTT	735				
Db	487 GCCCTGGCCCTGGTAAAGCTGGTGGATTGAAATCCAAAGAGAGCAACAGTTGTT	546				
QY	736 GCCAGATATGCTAAAGTGTACGTGGTGAATGAGTGTACAGGGAGGAGGGCTTCCC	795				
Db	547 GCCAGATATGCTAAAGTGTACGTGGTGAATGAGTGTACAGGGAGGAGGGCTTCCC	606				
QY	796 TGCAACAAGATCATCATGCTGTTGGGCCCTCGGTGGATGGAATGGGATTAACAGGAGT	854				
Db	607 TGCAACAAGATCATCATGCTGTTGGGCCCTCGGTGGATGGAATGGGATTAACAGGAGT	665				
RESULT 13						
ID	AAH70240/C					
XX	AAH70240 standard; cDNA; 516 BP.					
AC	AAH70240;					
DX	19-SEP-2001 (first entry)					
XX						
DE	Human cervical cancer marker nucleic acid 1514.					
XX						
KW	Cervical cancer; cytosstatic; pre-malignant condition; gene therapy; ss.					
OS	Homo sapiens.					
XX						
PN	WO200142467-A2.					
XX						
PD	14-JUN-2001.					
XX						
PF	08-DEC-2000; 2000WO-US33312.					
XX						

PR	08-DEC-1999;	99US-0169681.
PR	21-DEC-1999;	99US-0171350.
PR	14-MAR-2000;	2000US-0189315.
PR	12-MAY-2000;	2000US-0203791.
PR	09-JUN-2000;	2000US-0210600.
PR	21-JUL-2000;	2000US-0220114.
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
XX		
PI	Schlegel R, Deeds J, Berger A, Zhao X;	
XX		
DR	WPI: 2001-375006/39.	
XX		
PT	New isolated nucleic acid for diagnosing and treating cervical cancer	
PT	and for assessing and detecting compounds for treating the cancer -	
XX		
PS	Claim 1; Page 341; 1051pp; English.	
XX		
CC	The invention relates to novel genes (AAH68727-AAH73383) associated with	
CC	cervical cancer with cytostatic activity. The nucleic acids and encoded	
CC	polypeptides are useful: to assess if a patient is afflicted with	
CC	cervical cancer or has a pre-malignant condition; to monitor the	
CC	progression of cervical cancer or a premalignant condition in a patient;	
CC	and to select and/or assess the efficacy of a compound or therapy for	
CC	inhibiting cervical cancer in a patient. The nucleic acids may also be	
CC	useful for gene therapy.	
XX		
Q0	Sequence 516 BP; 124 A; 138 C; 108 G; 146 T; 0 other;	

Query Match	Similarity	14.3%	Score 463.2	DB 22	Length 516
Best Local	Similarity	99.2%	Pred. No. 1.1e-115		
Matches	476	Conservative	0	Mismatches	3
				Indels	1
				Gaps	
QY	379	TCAGAGACTGCTGCTCTTTGGAGAAAACTATAGTTGGCAAA-TTCCCATTAACCAACATGA	437		
Db	507	TCAGAGATTGGTGTCTTTGGAGAAAACTATAGTTGGCAAAATTCACATTAACCAACATGA	448		
QY	438	CTTCAGAAATTTTAAATAATATGACGCTGCTGTGTGAAGTCCACGAATTAAGTTGG	497		
Db	447	CTTCAGAAATTTTAAACAATAATGACGCTGCTGTGTGAAGTCCACGAATTAAGTTGG	388		
QY	498	CTGTATCTCTACCCGCTGCTCTCCAGTTCCAGGAAGGCAACACAGCAATCTCTGCAAGTGT	557		
Db	387	CTGTATCTCTACCCGCTGCTCTCTCCAGTTCCAGGAAGGCAACACAGCAATCTCTGCAAGTGT	328		
QY	558	CAGAAAAATGCTGCACTCTCAGATAGAGTATTCAGCTGTGGAAGATGAGACTTCAACACACA	617		
Db	327	CAGAAAAATGCTGCACTCTCAGATAGAGTATTCAGCTGTGGAAGATGAGACTTCAACACACA	268		
QY	618	TGCTTTGTGATGCTGTGTGTGAATGCAGGCCAATAAGATCTTCTGCAATGGGGAGGCTGTGC	677		
Db	267	TGCTTTGTGATGCTGTGTGTGAATGCAGGCCAATAAGATCTTCTGCAATGGGGAGGCTGTGC	208		
QY	678	CCTGGCCCTTGGTAAAGCTGTGTGATTTTAAATCCAAAGAAGAGCAAACTGTTTGTGC	737		
Db	207	CCTGGCCCTTGGTAAAGCTGTGTGATTTTAAATCCAAAGAAGAGCAAACTGTTTGTGC	148		
QY	738	CAGATATGTTAAGTGTCTGAGTGTGAGATAGCTGTCCAGGGAGGAGGCTTCCCTG	797		
Db	147	CAGATATGTTAAGTGTCTGAGTGTGAGATAGCTGTCCAGGGAGGAGGCTTCCCTG	88		
QY	798	CAAAACGATCATCCATGCTGTTGGGCTCTGGTGTGATGAGAAATGGGATTAACAGGAGATGAC	857		
Db	87	CAAAACGATCATCCATGCTGTTGGGCTCTGGTGTGATGAGAAATGGGATTAACAGGAGATGAC	28		
RESULT 14					
AAH71670/c					
ID	AAH71670 standard; CDNA; 481 BP.				
XX	AAH71670;				
XX	19-SEP-2001 (first entry)				

XX Human cervical cancer marker nucleic acid 2944.  
DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX Homo sapiens.  
XX WO200142467-A2.  
XX 14-JUN-2001.  
XX 08-DEC-2000; 2000WO-US33312.  
XX 08-DEC-1999; 99US-0169681.  
XX 21-DEC-1999; 99US-0171350.  
XX 14-MAR-2000; 2000US-0189315.  
XX 12-MAY-2000; 2000US-0203791.  
XX 09-JUN-2000; 2000US-0210600.  
XX 21-JUL-2000; 2000US-0220114.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Deeds J, Berger A, Zhao X;  
PI WPI; 2001-375006/39.  
XX New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX  
XX Claim 1; Page 597; 1051pp; English.  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
XX Sequence 481 BP; 116 A; 129 C; 102 G; 134 T; 0 other;  
XX  
Query Match 13.9%; Score 449.8; DB 22; Length 481;  
Best Local Similarity 99.4%; Pred. No. 4.7e-112;  
Matches 462; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 394 CTGGAGAACTATAGTTGGCAA-TCCTATTAACCAATGACTTCAAAATTTTAA 452  
DB 481 CTGGAGAACTATAGTTGGCAAATTTCCCATTAACCAATGACTTCAAAATTTTAA 422  
QY 453 AATATGAGCGCTGCTGTGAAGTCTCCAGATAGTTGGCTATCTACCT 512  
DB 421 CAATATATAGGCTGCTGTGAAGTCTCCAGATAGTTGGCTATCTACCT 362  
QY 513 GGTCTCTCAGTTAGGAAGCAACAGCAAAATCTCTGCAAGTTTTCAGAAAAATGCTGAC 572  
DB 361 GGTCTCTCAGTTAGGAAGCAACAGCAAAATCTCTGCAAGTTTTCAGAAAAATGCTGAC 302  
QY 573 TCCATGATAGTATAGTCTGGAAGATGACCTCAACACATGCTGTGATGCTGT 632  
DB 301 TCCATGATAGTATAGTCTGGAAGATGACCTCAACACATGCTGTGATGCTGT 242  
QY 633 GGTGAATCAGCCCAATGAAGATCTTCTGATGGGAGAGCCCTGGCCCTGGTAA 692  
DB 241 GGTGAATCAGCCCAATGAAGATCTTCTGATGGGAGAGCCCTGGCCCTGGTAA 182  
QY 693 AGCTGGTGAATTTGAATCAAGAAGAGCAAAAGTTTGTGACAGATATGTAAGT 752  
DB 181 AGCTGGTGAATTTGAATCAAGAAGAGCAAAAGTTTGTGACAGATATGTAAGT 122  
QY 753 GTCAGCTGTAGATAGTGTACAGGAGAGAGAGGCTCCCTGCAAAACAGATCATCA 812  
DB 121 GTCAGCTGTAGATAGTGTACAGGAGAGAGAGGCTCCCTGCAAAACAGATCATCA 62

QY 813 TGTCTGGCCCTGCGTATGATGAATGGGATAAACAGCATGTAC 857  
DB 61 TGTCTGGCCCTGCGTATGATGAATGGGATAAACAGCATGTAC 17  
RESULT 15  
AB056473/c  
ID AB056473 standard; cDNA; 668 BP.  
XX  
XX AC AB056473;  
XX DT  
XX 02-AUG-2002 (first entry)  
XX  
XX Human colon cancer related nucleotide sequence SEQ ID NO:168.  
XX  
XX DE  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200229086-A2.  
XX  
XX PD 11-APR-2002.  
XX  
XX PF 02-OCT-2001; 2001WO-US30732.  
XX  
XX PR 02-OCT-2000; 2000US-237271P.  
XX  
XX PA (FARB ) BAYER CORP.  
XX  
XX PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thlaglingam A, Lewis ME;  
XX WPI; 2002-426115/45.  
XX  
XX New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell  
PT or tissue type, and in antisense therapy -  
XX  
XX  
XX Claim 1; Fig 1; 796pp; English.  
XX  
XX AB056306 to AB060787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
CC encoded by the AB060776 to AB060787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (II) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence of  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists.  
XX  
XX Sequence 668 BP; 145 A; 152 C; 137 G; 175 T; 59 other;  
XX  
Query Match 13.1%; Score 424.2; DB 24; Length 668;  
Best Local Similarity 96.5%; Pred. No. 5.3e-105;  
Matches 462; Conservative 0; Mismatches 14; Indels 3; Gaps 3;  
QY 379 TCAGAGACTGCTGCTTGGAGAAACTATAGTTGGCAATTCCTATTAACCAATGAC 438  
DB 477 TCAAGATGAGTGTCTTGAAGAAACCTATAGTGGCAAAATTCCTATTAACCAATGAC 418  
QY 439 TTCAAAATTTTAAAAATATGACGCTGCTGTGAAGTCTCCAGATATGTTGGC 498  
DB 417 TTCAAAATTTTAAAAATATGAG-GTCAGCTGTGTAAAGTCTCCAGAAATAGTTGGC 359

QY 499 TGTATCTACCTGGTCTCTCCAGTTCAGGAAGGCACAAATCTGCAAGTGTTC 558  
|| |||||  
Db 358 TG-ATCTCTACCTGGTCTCTCCAGTTCAGGAAGGCACAAATCTGCAAGTGTTC 300  
|| |||||  
QY 559 AGAAAAAGCTGACTCCTAGGATAGAGTATTCAGTCTGGAAGATGACTCACCACACAT 618  
|| |||||  
Db 299 AG-AAAATGCTGACTCCTAGGATAGAGTATTCAGTCTGGAAGATGACTCACCACACAT 241  
|| |||||  
QY 619 GCTGTGATGCTGTGGATGAGCAATGAAAGATCTTCTGCATGGGGAGGCTTGCC 678  
|| |||||  
Db 240 GCTGTGATGCTGTGGATGAGCAATGAAAGATCTTCTGCATGGGGAGGCTTGCC 181  
|| |||||  
QY 679 CTGGCCCTGTTAAAGCTGTGATTTGAATCCAAAGAGACAAACAGTTTGTGCC 738  
|| |||||  
Db 180 CTGGCCCTGTTAAAGCTGTGATTTGAATCCAAAGAGAGACAAACAGTTTGTGCC 121  
|| |||||  
QY 739 AGATATGTTAAAGTGTGAGCTGTGAGATAGCTGTCAAGGAGCAGGAGGCTTCCCTGC 798  
|| |||||  
Db 120 AGATATGTTAAAGTGTGAGCTGTGAGATAGCTGTCAAGGAGCAGGAGGCTTCCCTGC 61  
|| |||||  
QY 799 AAACAGATCATGCTGTGGGCTCGGTGATGGAATGGGATAAACAGGATGTAC 857  
|| |||||  
Db 60 AAACAGATCATGCTGTGGGCTCGGTGATGGAATGGGATAAACAGGATGTAC 2  
|| |||||

Search completed: February 7, 2003, 02:08:44  
Job time : 452 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2003, 04:00:54 : Search time 112 Seconds  
(Without alignments)  
13724.926 Million cell updates/sec

Title: US-09-830-762-1  
Perfect score: 3243  
Sequence: 1 ggcctcgcgttcctcgtggtg.....gtcgtatataaaaaaaaaa 3243

Scoring table:  
IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 408267 seqs, 237001491 residues

Total number of hits satisfying chosen parameters: 816534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCF\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2842.6	87.7	3016	10 US-09-882-529-1	Sequence 1, Appl
2	2842.6	87.7	3016	10 US-09-882-529-1	Sequence 1, Appl
3	299	9.2	299	10 US-09-920-300A-946	Sequence 946, App
4	284	8.2	299	10 US-10-033-528-946	Sequence 946, App
5	246.8	7.6	284	10 US-09-962-436-149	Sequence 149, App
6	226	7.0	238	10 US-09-964-824A-308	Sequence 308, App
7	202.8	6.3	214	9 US-09-796-692-7357	Sequence 7357, App
8	202.8	6.3	214	10 US-09-604-287A-24	Sequence 24, Appl
9	202.8	6.3	214	10 US-09-339-338-24	Sequence 24, Appl
10	202.8	6.3	214	10 US-10-007-805-24	Sequence 24, Appl
11	180	5.6	180	9 US-10-046-935-1590	Sequence 1590, App
12	140.4	4.3	2762	10 US-09-957-635-1	Sequence 1590, App
13	140.4	4.3	2217	10 US-09-957-635-3	Sequence 1, Appl
14	121.2	3.7	3348	12 US-10-044-090-664	Sequence 664, App
15	121.2	3.7	3348	12 US-10-044-090-664	Sequence 664, App
16	70.2	2.2	71	10 US-09-998-598-2276	Sequence 2276, App
17	44.8	1.4	533	10 US-10-015-219-138	Sequence 138, App
18	44.8	1.4	533	10 US-09-777-564-138	Sequence 138, App
19	44.8	1.4	649	9 US-10-015-219-375	Sequence 375, App

20	44.8	1.4	649	10 US-09-777-564-375	Sequence 375, App
21	42.4	1.3	1192	10 US-09-864-761-10189	Sequence 10189, A
22	41	1.3	1136	9 US-09-938-842A-3287	Sequence 3287, App
23	40	1.2	127197	9 US-09-754-853A-1	Sequence 1, Appl
24	39.6	1.2	408	10 US-09-960-352-6263	Sequence 6263, App
25	39.6	1.2	2000	9 US-09-938-842A-4600	Sequence 4600, App
26	39.4	1.2	229	10 US-09-969-373-1347	Sequence 1347, App
27	39.4	1.2	302250	10 US-09-962-832-154	Sequence 154, App
28	39.4	1.2	640681	10 US-09-790-988-1	Sequence 1, Appl
29	39.2	1.2	2672	12 US-10-044-090-699	Sequence 699, App
30	39	1.2	650	10 US-09-879-536-189	Sequence 189, App
31	39	1.2	513509	9 US-09-754-853A-4	Sequence 4, Appl
32	38.4	1.2	601	10 US-09-820-002-13	Sequence 13, Appl
33	38.4	1.2	2073	10 US-09-815-242-8362	Sequence 8362, App
34	38.4	1.2	2109	10 US-09-815-242-4555	Sequence 4555, App
35	38.4	1.2	21784	10 US-09-820-002-3	Sequence 3, Appl
36	38.2	1.2	640681	10 US-09-790-988-1	Sequence 1, Appl
37	38	1.2	237	10 US-09-880-107-11727	Sequence 10804, A
38	38	1.2	435	10 US-09-864-761-20804	Sequence 20804, A
39	38	1.2	447	10 US-09-864-761-4048	Sequence 4048, App
40	38	1.2	1089	10 US-09-925-302-301	Sequence 301, App
41	38	1.2	2184	9 US-09-880-192-31	Sequence 31, Appl
42	37.8	1.2	401	9 US-09-946-607-678	Sequence 678, App
43	37.8	1.2	401	10 US-09-795-668-678	Sequence 678, App
44	37.8	1.2	401	10 US-09-795-668-678	Sequence 678, App
45	37.8	1.2	406	10 US-09-960-352-10265	Sequence 10265, A

#### ALIGNMENTS

RESULT 1  
US-09-882-529-1  
Sequence 1, Application US/09882529  
Patent No. US20020132317A1  
GENERAL INFORMATION:  
APPLICANT: Peyman, John A  
APPLICANT: da Silva, Antonio  
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
FILE REFERENCE: 15966-771  
CURRENT APPLICATION NUMBER: US/09/882, 529  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 60/211, 565  
PRIOR FILING DATE: 2000-06-15  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 3016  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (123)..(2579)  
US-09-882-529-1  
Query Match 87.7%; Score 2842.6; DB 10; Length 3016;  
Best Local Similarity 96.3%; Pred. No. 0;  
Matches 2963; Conservative 0; Mismatches 9; Indels 105; Gaps 1;

QY 151 CATATCTGGAACATCACTATGCTTTGAAGCCGAAAGGAAATTAACATTAAAGCT 210  
DB 45 CAGGCTGGAACATCACTATGCTTTGAAGCCGAAAGGAAATTAACATTAAAGCT 104  
QY 211 CCCCCGGGACCTGAGAGATGAGCTTTTCATGTGCGCGAGCAGCAGCTTAAATGAA 270  
DB 105 CCCCCGGGACCTGAGAGATGAGCTTTTCATGTGCGCGAGCAGCAGCTTAAATGAA 164  
QY 271 AATTCAGTAGGATTAACCTGCTCTCACTCTGTTTCAGAAAGTCTTGCAGATCTTT 330  
DB 165 AAA----- 167

Oy 331 CCTCAGTGGAGAAAGGGAATACAGAGAATGTCTCCCTACAGTGTCTCAGAGACTGCT 390  
Db 168 -----TCAGAGCTGGT 179  
Oy 391 GCTCTTGGAGAAACTATAGTTGGCAATTCCTCCATTACCCACAATGACTTCAAAATTTTA 450  
Db 180 GCTCTTGGAGAAACTATAGTTGGCAAAATTCCTATTAACCAATGACTTCAAAATTTTA 239  
Oy 451 AAAAATAAGAGCGTCAGCTGTGTGAAGTCTCCAGAAATAGTTGGCTGTATCTCTACC 510  
Db 240 AAAAATAAGAGCGTCAGCTGTGTGAAGTCTCCAGAAATAGTTGGCTGTATCTCTACC 299  
Oy 511 CTGGTCTCTCAGTTCCAGGAAGCAACAGCAATCTCTCAGTGTTCAGAAAAATGCTG 570  
Db 300 CTGGTCTCTCAGTTCCAGGAAGCAACCAATCTCTCAGTGTTCAGAAAAATGCTG 359  
Oy 571 ACTCCTAGATAGAGTTATCAGTCTGGAAAAATGACCTCACACACATCTGTTGATGCT 630  
Db 360 ACTCCTAGATAGAGTTATCAGTCTGGAAAAATGACCTCACACACATCTGTTGATGCT 419  
Oy 631 GTGGTGAATGACGCGCAATGAGATCTTCGATGGGGGGGCGCTGGCGCTGGTGA 690  
Db 420 GTGGTGAATGACGCGCAATGAGATCTTCGATGGGGGGGCGCTGGCGCTGGTGA 479  
Oy 691 AAAGCTGTGATTTGAAAAATCCAGAGAGAGCAAAACAGTTTGTGCCAGATATGTTAA 750  
Db 480 AAAGCTGTGATTTGAAAAATCCAGAGAGAGCAAAACAGTTTGTGCCAGATATGTTAA 539  
Oy 751 GTGTGAGCTGTGATAGTACGTCTCACGGGAGCAGGAGGCTTCCCTGCAGACATATC 810  
Db 540 GTGTGAGCTGTGATAGTACGTCTCACGGGAGCAGGAGGCTTCCCTGCAGACATATC 599  
Oy 811 CATGCGTTGGGGCGCTGGTGGATGGAATGGGATTAACAGGAGTGTCTGGAAAACTGCG 870  
Db 600 CATGCGTTGGGGCGCTGGTGGATGGAATGGGATTAACAGGAGTGTCTGGAAAACTGCG 659  
Oy 871 AGGGCATTGTAGTATCTGATTAATGATCATCTATAAAAATACACATTAAAGACATTA 930  
Db 660 AGGGCATTGTAGTATCTGATTAATGATCATCTATAAAAATACACATTAAAGACATTA 719  
Oy 931 GCAATTCACGCTTGAGCTCTGGGATTTTTCAGTTCCCTCTGAATTTGTGTACAAAGACT 990  
Db 720 GCAATTCACGCTTGAGCTCTGGGATTTTTCAGTTCCCTCTGAATTTGTGTACAAAGACT 779  
Oy 991 ATTGTAGACTATCCGGGTAGTTTGAAGGAAAGCCATGATGAGTAAATTTGAAGAA 1050  
Db 780 ATTGTAGACTATCCGGGTAGTTTGAAGGAAAGCCATGATGAGTAAATTTGAAGAA 839  
Oy 1051 ATTCACCTGTGAGCAATGAGAACCCCTACTGTTGCTGCCCTTTAAAGCTCTTCAGAAATTC 1110  
Db 840 ATTCACCTGTGAGCAATGAGAACCCCTACTGTTGCTGCCCTTTAAAGCTCTTCAGAAATTC 899  
Oy 1111 ATTCCTAGGGAAGAGTACGTGGGACAAAGAACCCCTCTTCTTCAATGCATGGTCG 1170  
Db 900 ATTCCTAGGGAAGAGTACGTGGGACAAAGAACCCCTCTTCTTCAATGCATGGTCG 959  
Oy 1171 AACCACTGACCTCCAGATTGTCAGGGCCACATGAAATGGCAGAGCAGATGTAAT 1230  
Db 960 AACCACTGACCTCCAGATTGTCAGGGCCACATGAAATGGCAGAGCAGATGTAAT 1019  
Oy 1231 GTTAAATCTGTAAAACCACATGATATTTACAGTTGACCTGTGGCAAAAGTCAATTTCA 1290  
Db 1080 CAAGCAGGAGTTGAATGGAATTCGAAATTTCTGCAAAAGGCTTAAACAGTTTCAACG 1139  
Oy 1291 CAAGCAGGAGTTGAATGGAATTCGAAATTTCTGCAAAAGGCTTAAACAGTTTCAACG 1350  
Db 1351 TCCCAATTTGTTACTGTCACAAAGAGTTTAACTTCTCTGTAATATATATACCATGTA 1410  
Oy 1410 TCCCAATTTGTTACTGTCACAAAGAGTTTAACTTCTCTGTAATATATATACCATGTA 1199  
Oy 1411 CTGTGCAATTCAGAAATTTCTTAAACCTCAGATATTTAAACATGCATGAAGAGTGTGG 1470

Db 1200 CTGTGCAATTCAGAAATTTCTTAAACCTCAGATATTTAAACATGATGAAGAGTGTGG 1259  
Oy 1471 GAAAAATGATTTGACCAAAATATTAATCTTCATTTCTCTTCCGCTGGGCTGGGAAC 1530  
Db 1260 GAAAAATGATTTGACCAAAATATTAATCTTCATTTCTCTTCCGCTGGGCTGGGAAC 1319  
Oy 1531 ATGGAATTAAGAGAAAGAAACACAGCAGAGATTTGTTGATGAAGTAAATTAATTTGCC 1590  
Db 1320 ATGGAATTAAGAGAAAGAAACACAGCAGAGATTTGTTGATGAAGTAAATTAATTTGCC 1379  
Oy 1591 AAAGACATGTAACACACAGTTAACTGTAAATTTTGATCTTTTCAACAGATTTGGAG 1650  
Db 1380 AAAGACATGTAACACACAGTTAACTGTAAATTTTGATCTTTTCAACAGATTTGGAG 1439  
Oy 1651 ATATATAGGCTTTGAGTTCTGAATTTGGCAAGAGTCCAAATGCTGATTTGAACAT 1710  
Db 1440 ATATATAGGCTTTGAGTTCTGAATTTGGCAAGAGTCCAAATGCTGATTTGAACAT 1499  
Oy 1711 TACAGTGTCCCGAGTCAACAGAGAGAGAAAGAAATGGGCTTGAAGCTAGATCT 1770  
Db 1500 TACAGTGTCCCGAGTCAACAGAGAGAGAAAGAAATGGGCTTGAAGCTAGATCT 1559  
Oy 1771 CCTGCCATCAATCTGATGGGATTTCAACGTGGAAGATGTATAGGCGCCAGCATGATC 1830  
Db 1560 CCTGCCATCAATCTGATGGGATTTCAACGTGGAAGATGTATAGGCGCCAGCATGATC 1619  
Oy 1831 CAAAGATCTGATGTCCTCAGAACCAACCAATCATTTGAGAAATATCTATTTCTGATCT 1890  
Db 1620 CAAAGATCTGATGTCCTCAGAACCAACCAATCATTTGAGAAATATCTATTTCTGATCT 1679  
Oy 1891 GGGAGAAAGAACATGACATTTTGTCTCAGCTTCAGAAACTTCAAGTGTCTCATACA 1950  
Db 1680 GGGAGAAAGAACATGACATTTTGTCTCAGCTTCAGAAACTTCAAGTGTCTCATACA 1739  
Oy 1951 GAAATTAATCAGCCACAGGAAGACAGATTTAGAGATTTGAAGAGCCCGGCTGACCTCAT 2010  
Db 1740 GAAATTAATCAGCCACAGGAAGACAGATTTAGAGATTTGAAGAGCCCGGCTGACCTCAT 1799  
Oy 2011 GAGTGGTATGATGACATTTGAATGATGCTTTGAAAGTACAGAGGAAATGSCAAGGAA 2070  
Db 1800 GAGTGGTATGATGACATTTGAATGATGCTTTGAAAGTACAGAGGAAATGSCAAGGAA 1859  
Oy 2071 AAGGACGAGGCGCTTGGGCTGCTGTAGACAGTGGACTATTCACGACCAAAACCCAA 2130  
Db 1860 AAGGACGAGGCGCTTGGGCTGCTGTAGACAGTGGACTATTCACGACCAAAACCCAA 1919  
Oy 2131 GACGAATGAAGAAATATCATATTTCTGAATGTCTGTGCTCCAACTCAAGACTT 2190  
Db 1920 GACGAATGAAGAAATATCATATTTCTGAATGTCTGTGCTCCAACTCAAGACTT 1979  
Oy 2191 CTGAGTCAAAAGAAACAGTTGAAAAATGTGTTGGCAGCTTCTAAAGTGGAGAAATTA 2250  
Db 1980 CTGAGTCAAAAGAAACAGTTGAAAAATGTGTTGGCAGCTTCTAAAGTGGAGAAATTA 2039  
Oy 2251 GACAATGAGGCTCTTATGCTGCTCTTTCAAAGAAAGAAAGAAATGATGAGAAAAACTG 2310  
Db 2040 GACAATGAGGCTCTTATGCTGCTCTTTCAAAGAAAGAAAGAAATGATGAGAAAAACTG 2099  
Oy 2311 CACAGGCAACCTGTGAGCCATAGGCTGTTTCAGCAAGTCCCATACAGTTCTGCAATGTG 2370  
Db 2100 CACAGGCAACCTGTGAGCCATAGGCTGTTTCAGCAAGTCCCATACAGTTCTGCAATGTG 2159  
Oy 2371 GTATGAGAGTGTGGCTTCAAGAAATGTCTGACACCTTTGGGATGCCAAATATAGGAGCT 2430  
Db 2160 GTATGAGAGTGTGGCTTCAAGAAATGTCTGACACCTTTGGGATGCCAAATATAGGAGCT 2219  
Oy 2431 GGCATATACCTTACCAAGAACCTCAAAAACTGGGAGAGAGGCAAGAAAAATCTCTGCT 2490  
Db 2220 GGCATATACCTTACCAAGAACCTCAAAAACTGGGAGAGAGGCAAGAAAAATCTCTGCT 2279  
Oy 2491 GCAGATTAAGCTGATGATGTGTTGAGGCTGAAGTATCTCAGAGGCTCTTGTGCCAGGGA 2550



US-09-962-436-149/c  
; Sequence 149, Application US/09962436  
; Patent No. US20020081301A1  
; GENERAL INFORMATION:  
; APPLICANT: Soppec, Daniel  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu  
; FILE OF INVENTION: Sets  
; FILE REFERENCE: 689290-75  
; CURRENT APPLICATION NUMBER: US/09/962,436  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,082  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/234,924  
; PRIOR FILING DATE: 2000-09-25  
; NUMBER OF SEQ ID NOS: 568  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 149  
; LENGTH: 284  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-962-436-149

Query Match 8.8%; Score 284; DB 10; Length 284;  
Best Local Similarity 100.0%; Pred. No. 5.4e-64;  
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2951 GTGCTTATATCTGTCTTATGTCCTGGGGTGGGAGTAGATACCAATGAACACTT 3010  
DB 284 GTGCTTATATCTGTCTTATGTCCTGGGGTGGGAGTAGATACCAATGAACACTT 225  
QY 3011 TCAGGACCTTCCTCTCTTCGAGTTGTTCTTAATCTCCCTTACTAGAGAGATAATA 3070  
DB 224 TCAGGACCTTCCTCTCTTCGAGTTGTTCTTAATCTCCCTTACTAGAGAGATAATA 165  
QY 3071 TTTTCATATATGAAGAAATTTTCTAGTATATACGAGCCCTTTATTTCTAAAT 3130  
DB 164 TTTTCATATATGAAGAAATTTTCTAGTATATACGAGCCCTTTATTTCTAAAT 105  
QY 3131 GATGATAGTATAAATGTTAGATACAGATGATTTAGATTTCCAGAAATATTAT 3190  
DB 104 GATGATAGTATAAATGTTAGATACAGATGATTTAGATTTCCAGAAATATTAT 45  
QY 3191 AAGTCTTACGATGAATAATCATCTTGTCTGATTAA 3234  
DB 44 AAGTCTTACGATGAATAATCATCTTGTCTGATTAA 1

RESULT 5  
US-09-964-824A-308  
; Sequence 308, Application US/09964824A  
; Patent No. US20020102531A1  
; GENERAL INFORMATION:  
; APPLICANT: Horrigan, Stephen  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu  
; FILE OF INVENTION: Sets  
; FILE REFERENCE: 689290-73  
; CURRENT APPLICATION NUMBER: US/09/964,824A  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US/60/236,033  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US/60/236,032  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US/60/236,028  
; PRIOR FILING DATE: 2000-09-28  
; NUMBER OF SEQ ID NOS: 583  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 308  
; LENGTH: 260  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-964-824A-308

Query Match 7.6%; Score 246.8; DB 10; Length 260;

Best Local Similarity 99.2%; Pred. No. 2.3e-54;  
Matches 248; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2824 CTTACCTGGGTGAGACTAACCAATATGACCATGAGGCTCAAGAGTGCTTGATA 2883  
DB 1 CTTACCTGGGTGAGACTAACCAATATGACCATGAGGCTCAAGAGTGCTTGATA 60  
QY 2884 TATCCATGAGTTATCTGATGAGTGAAGGCTGATTTGAAAGGACTAGCCACATAG 2943  
DB 61 TATCCATGAGTTATCTGATGAGTGAAGGCTGATTTGAAAGGACTAGCCACATAG 120  
QY 2944 CATCTTAGGCTTTATCTGTTTATGCTTGGGGTGGGTAGTATACCAATGA 3003  
DB 121 CATCTTAGGCTTTATCTGTTTATGCTTGGGGTGGGTAGTATACCAATGA 180  
QY 3004 AACCTTCAGGACCTTCCTCTTCGAGTTGTTCTTATCTCCTTACTAGAGAG 3063  
DB 181 AACCTTCAGGACCTTCCTCTTCGAGTTGTTCTTATCTCCTTACTAGAGAG 240  
QY 3064 ATAAATATT 3073  
DB 241 ATAAATATT 250

RESULT 6  
US-09-796-692-7357  
; Sequence 7357, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
; FILE REFERENCE: 2077,001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 7357  
; LENGTH: 238  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-796-692-7357

Query Match 7.0%; Score 226; DB 9; Length 238;  
Best Local Similarity 99.6%; Pred. No. 5.3e-49;  
Matches 237; Conservative 0; Mismatches 0; Indels 1; Gaps 1;



```
RESULT 10
US-10-007-805-24/c
; Sequence 24, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepier, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-007-805-24

Query Match          6.3%; Score 202.8; DB 12; Length 214;
Best Local Similarity 96.7%; Pred. No. 5.2e-43;
Matches 207; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2593 ATAGATGTCATGACAGTGTGGTGCATGTCCTCCAGCCCGAAACCTTTTATTTT 2652
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 214 AAAGATGGTCATGACAGTGTGGTGCATGTTCCACCCGAAACCTTTTATTTT 155
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2653 ACTGCATGACAGCTATACCTCAGTATTTGTGACATGACCCAGCAATATGTACAGTCA 2712
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 154 ACTGCATGACAGCTATACCTCAGTATTTGTGACATGACCCAGCAATATGTACAGTCA 95
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2713 CAAGATTACTCATGACAGACCAATGAGACCCCTTTGACACAGCATCCTTTGGAGGGATTGCA 2772
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 94 CAAGATTACTCATGACAGACCAATGAGACCCCTTTGACACAGCATCCTTTGGAGGGATTGCA 35
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2773 AGTGGAGCCCTGTGATTAATCTCTACATCAT 2806
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 34 AGTGGAGCCCTGTGATTAATTTTACATCAT 1

RESULT 11
US-10-046-935-1590
; Sequence 1590, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aljun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1590
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-1590
```

```
Query Match          5.6%; Score 180; DB 9; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.8e-37;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2526 ACTCAGAGCTTCTCTGCGAGGACATCCGTTAATATATGTTCCCAACCACTGATCC 2585
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 ACTCAGAGCTTCTCTGCGAGGACATCCGTTAATATATGTTCCCAACCACTGATCC 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2586 TGGACCTATAGATGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAACCTTTGT 2645
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 TGGACCTATAGATGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAACCTTTGT 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2646 TATTTTATGTCATGACAGGCTATACCTCAGTATTTGTGACATGACCCAGCAATATGT 2705
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 TATTTTATGTCATGACAGGCTATACCTCAGTATTTGTGACATGACCCAGCAATATGT 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-878-178-1590
; Sequence 1590, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1590
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-878-178-1590

Query Match          5.6%; Score 180; DB 9; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.8e-37;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2526 ACTCAGAGCTTCTCTGCGAGGACATCCGTTAATATATGTTCCCAACCACTGATCC 2585
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 ACTCAGAGCTTCTCTGCGAGGACATCCGTTAATATATGTTCCCAACCACTGATCC 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2586 TGGACCTATAGATGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAACCTTTGT 2645
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 TGGACCTATAGATGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAACCTTTGT 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2646 TATTTTATGTCATGACAGGCTATACCTCAGTATTTGTGACATGACCCAGCAATATGT 2705
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 TATTTTATGTCATGACAGGCTATACCTCAGTATTTGTGACATGACCCAGCAATATGT 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-957-635-1/c
; Sequence 1, Application US/09957635
; Patent No. US20020120112A1
; GENERAL INFORMATION:
; APPLICANT: Shipp, Margaret
; APPLICANT: Aguilar, Ricardo
; APPLICANT: Gu, Liqun
; APPLICANT: Takeyama, Kunihiko
; TITLE OF INVENTION: LYMPHOMA ASSOCIATED MOLECULES AND USES THEREFOR
; FILE REFERENCE: DFN-036
; CURRENT APPLICATION NUMBER: US/09/957,635
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/233,791
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
```

```

; SEQ ID NO 1
; LENGTH: 2762
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (167)..(2383)
;
US-09-957-635-1

```

Query Match	4.38;	Score 140.4;	DB 10;	Length 2762;
Best Local Similarity	93.08;	Pred. No. 3.3e-26;		
Matches 147; Conservative	0;	Mismatches 11;	Indels 0;	

<b>QY</b>	1	GGGCTTCGTGTTCCCTGGGGTGTGAACCGTGACATCCGCCGCCGCAGAGCTTTAGAGCTC	60
<b>Db</b>	317	GGGCTTCGTGTTCCCTGGGGTGTGAACCGTGACATCCGCCGCCGCAGAGACTTTAAGGCTC	258
<b>QY</b>	61	GGAAGTAGCTTCCAGCACTTCCTTCGTACTCGGGGGGCCGAGATTGTACACCGCACGAGGA	120
<b>Db</b>	257	GGAAGTAGCTTCCAGCACTTCCTTCGTACTCGGGGGGCCGAGATTGTACACCOCGACGAGGA	198
<b>QY</b>	121	GCGGGAGCGCGCGCGCAGAAGTGGGCGACCATATCTG	158
<b>Db</b>	197	GCGGGAGCGCGCGCGCGCAGAGTGGGCGACCATGCGCTCTG	160

```

RESULT 14
US-09-957-635-3/c
Sequence 3, Application US/09957635
Patent No. US20020120112A1
GENERAL INFORMATION:
APPLICANT: Shipp, Margaret
APPLICANT: Aguilar, Ricardo
APPLICANT: Gu, Hiquin
APPLICANT: Takeyama, Kunihiko
TITLE OF INVENTION: LYMPHOMA ASSOCIATED MOLECULES AND USES THEREFOR
FILE REFERENCE: DFN-036
CURRENT APPLICATION NUMBER: US/09/957,635
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/233,791
PRIOR FILING DATE: 2000-09-19
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2217
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2217)
US-09-957-635-3

```

Query Match	Similarity	4.3%	Score 140	DB 10	Length 2217
Best Local	Similarity	96.6%	Pred. M.3.7e-26		
Matches 143	Conservative	0	Mismatches 5	Indels 0	Gaps 0
QY	1	GGGCTTCGTGTCCCGGGGCTGTACCCGTCACCTCCCGCGCCGAGACTTAGACCTCT	60		
Db	151	GGGGCTTCGTGTCCCGGGGCTGTACCCGTCACCTCCCGCGCCGAGACTTAGACCTCT	92		
QY	61	GGAGTAGCTCTCCAGCTTCTCTGTACTCGGGGGCCGAGCTTGTACACCCGACAGAGA	120		
Db	91	GGAAATAGCTCTCCAGCTTCTCTGTACTCGGGGGCCGAGCTTGTACACCCGACAGAGA	32		
QY	121	GGGGGACCGCGCGGGCGCAGAAAGTGGGCC	148		
Db	31	GGGGGACCGCGCGGGCGCAGAGTGGAGGC	4		

RESULT 15  
US-10-044-090-664  
; Sequence 664, Application US/10044090  
; Patent No. US20020137081A1

```

? GENERAL INFORMATION:
? APPLICANT: Olga Bandman
? TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
? FILE REFERENCE: PA-0028 US
? CURRENT APPLICATION NUMBER: US/10/044, 090
? CURRENT FILING DATE: 2002-01-09
? NUMBER OF SEQ. ID NOS: 850
? SOFTWARE: PERL Program
? SEQ ID NO 664
? LENGTH: 3348
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Incyte ID No. US20020131081A1 475203.8
US-10-044-090-664

Query Match          3.7%; Score 121.2; DB 12; Length 3348;
Best Local Similarity 54.1%; Pred. No. 3.5e-21;
Matches 293; Conservative 0; Mismatches 243; Indels 6; Gaps 2;

```

Query Match	3.7%	Score 121.2	DB 12	Length 3348
Best Local Similarity	54.1%	Pred. No. 3.5e-21		
Matches 293	Conservative 0	Mismatches 243	Indels 6	Gaps 2
QY	560	GAAAAATGCTGACTCTCTAGAGATAGATTATCATGTCGTGAAAGATGACCTCACACACATG	619	
Db	428	GGACAGCTCTTGAGCCCTGGCGTTGTGATGTGTGCAGACAGGGATGACCTTGACACGGCTTC	487	
QY	620	CTGTGTGATGCTGTGTGTGAATGAGCCCATGAGATCTTCTGTATGGGGAGGAGCCCTGGGCC	679	
Db	488	CTGTGTGATGTGTGTGTGATGATGCATCTAATGAGAGACCTTAAGCATTTATGGTGTGCGCTGGCCG	547	
QY	680	TGGCCCTGGTAAAGCTGGTGGATTTGAAATTCAGAAAGAGAGCAAACTGTTGTTGCCA	739	
Db	548	CTGGCCTCTCAAAAGACAGCTGGCCCTGAGCCTCCAGGCGGACCTGTGACCGATAGTGAAAG	607	
QY	740	GATATGTGTAAGTGTACAGCTGGTGTGAGATAGCTGTCAAGGGAGACAGGAGGCTTCCCTGCA	799	
Db	608	GAGAGGGACAGACTCTTACCGGGCAATGCCCATCTCCAAAGGACAGAAAGCTGCCCCTACC	667	
QY	800	AACACATCATCCATCTGTGTGGCCCTCGGTGATGAAATGGATTAACAGGAGATGACTG	859	
Db	668	ACCAGGTGATCATGACAGTGGGGCCCCGGTGGAGGGATATGAGGCCCCGAGATGTGTGT	727	
QY	860	GAAAGCTGCAGAGGGCCATTTGTAAGTATTTCTAATTATGTCAATTAATAAATACTACACA	919	
Db	728	ACCTATTAAAGGAGACTGTGCAACTCAGTC-----TCTGTCTAGGCCGAAATAATACAAAGTA	782	
QY	920	TTAAGACAGTAGCAATTCACGCTTGAGCTGTGGATTTTTCAGTTCCCTCGAATTTCT	979	
Db	783	CCGATATCCA-TACCATCCCTCGCTATTAGTTCGTGAGTCTTTGGCTTTCCCTTAGGCCGAT	841	
QY	980	GTAACAAGACTATTTGAGACTATTCGCGGTTAGTTTGCAGAGGAAGCCAAATGATGAGTA	1039	
Db	842	GGTGGAGCCATTTTCTGTGCTCATAGAGAAACTTCCAAATTCAGAAAGAGATGACACT	901	
QY	1040	ATTGGAAGAAATTCACCGGTGAGACAATGAGAGACCCCTGCTGTGCGCTTAAAGCTG	1099	
Db	902	CGTTGAAGAAATCTAATCTTGTGATGTATCTGAGAGACACTGTTGAGGCCCTTTGCAGAAG	961	
QY	1100	CT 1101		
Db	962	CT 963		

Search completed: February 7, 2003, 07:01:09  
 Job time : 140 secs

Search completed: February 7, 2003, 07:01:09  
Job time : 140 secs

Search completed: February 7, 2003, 07:01:09  
Job time : 140 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2003, 01:59:24 ; Search time 2874 Seconds  
(without alignments)  
18274.866 Million cell updates/sec

Title: US-09-830-762-1  
Perfect score: 3243  
Sequence: 1 gggcttcgtcttcctcgtggtc.....gtctgattaaaaaaaaa 3243

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2052	63.3	2871	11	BC017463 Homo sapi
2	973	30.0	993	9	AL568673 AL568673
3	922	28.4	965	9	AL540976 AL540976
4	733.4	22.6	852	12	BG680130 BG680130
5	713.6	22.0	724	14	BQ014467 BQ014467
6	705.8	21.8	742	12	BG540093 BG540093

7	659.6	20.3	901	14	BQ923651 AGENCOURT
8	649.6	20.0	791	12	BG284006 BG284006
c	646.6	19.9	679	13	BM664435 UI-E-CL1
c	645	19.9	658	14	BQ028103 UI-H-CO0
11	604.8	18.6	608	14	BM829571 K-EST0102
12	600.4	18.5	705	13	B1857151 603383747
c	586.8	18.1	601	10	AM851285 IL3-CT022
13	586.2	18.1	732	13	B1836699 603089634
15	584.2	18.0	610	10	AM851000 IL3-CT022
c	584	18.0	587	9	A1307751 t224f11.x
16	579.4	17.9	582	10	AM850996 IL3-CT022
c	577.4	17.8	579	10	AM851276 IL3-CT022
c	567.4	17.5	569	13	BM503214 1973b06.x
c	565.8	17.4	811	12	BG202125 RST21479
c	558.4	17.2	568	14	BQ268851 1k21d03.x
c	551.8	17.0	605	9	AA151346 z125b01.f
c	551	17.0	551	12	BF111184 7n4d612.x
c	550	17.0	550	9	A1860767 w105d09.x
c	546	16.8	706	10	BE612556 601452069
c	544.8	16.8	548	10	BE326805 hr65b06.x
c	544.8	16.8	781	12	BE962693 601656069
c	542	16.7	554	12	BG768000 602743759
c	538	16.6	765	13	B1855784 603382722
c	532.8	16.4	551	9	A1670955 w012c08.x
c	520	16.0	520	13	BM503372 1973b06.y
c	512	15.8	572	14	BQ362709 MR2-ST022
c	509.2	15.7	529	9	A1670956 w012c09.x
c	507	15.6	507	10	AM474917 xy21f12.x
c	500	15.4	500	9	A1129360 qc34a12.x
c	497.4	15.3	527	12	BF057083 7k15f01.x
c	497.2	15.3	542	13	B1010748 MR2-EN009
c	489	15.1	489	9	A1304947 q020b12.x
c	487	15.0	487	9	AA431590 zw70b08.f
c	487	15.0	615	14	BQ269118 1k21d03.y
c	485	15.0	510	14	NS4275 yv66c11.sl
c	481	14.8	481	9	A1653467 t94f102.x
c	481	14.8	704	12	BF684485 602142974
c	480.8	14.8	651	9	AL602483 DKP2p686P
c	480	14.8	480	9	A1523805 t996c06.x

## ALIGNMENTS

RESULT 1  
BC017463 2871 bp mRNA linear HTC 13-NOV-2001  
LOCUS  
DEFINITION Homo sapiens, clone IMAGE:4862464, mRNA.  
ACCESSION BC017463  
VERSION BC017463.1 GI:16907195  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 2871)  
Strausberg, R.  
Direct Submission  
Submitted (13-NOV-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DNP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
Info@cgsc.bc.ca  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,

REMARK  
COMMENT



Db 1837 CATATCTTACCTTGGGAGAAAGAACATGACATTTTGTCTCAGCTTCAGAAAACTTCA 1896  
QY 1936 AGTGCTCCATCAGAAATATATCAGCCAGAGAGACAGATAGACATTTGAAGAGACC 1995  
Db 1887 AGTGTCTCCATCAGAAATATATCAGCCAGAGAGACAGATAGACATTTGAAGAGACC 1956  
QY 1996 CGGGCTGACCTCATTTGAGGTGTTATGAACATTTGAAGATTTGTTAAAGTACAGAG 2055  
Db 1957 CGGGCTGACCTCATTTGAGGTGTTATGAACATTTGAAGATTTGTTAAAGTACAGAG 2016  
QY 2056 GAAATGCGAAGAAAGAGAGCGCTTTGGCGCTGTTAGACAGTGCATTTACG 2115  
Db 2017 GAAATGCGAAGAAAGAGAGCGCTTTGGCGCTGTTAGACAGTGCATTTACG 2076  
QY 2116 CAACAAAAAACCCAGAGCAATGAAGAAATATCATTTTCTGAATTTGCTTGGCT 2175  
Db 2077 CAACAAAAAACCCAGAGCAATGAAGAAATATCATTTTCTGAATTTGCTTGGCT 2136  
QY 2176 CCAACTCAGAGCTTCTGATCAAAAGAACAGTTTGAATAATGGTTTGCAGGTTCTA 2235  
Db 2137 CCAACTCAGAGCTTCTGATCAAAAGAACAGTTTGAATAATGGTTTGCAGGTTCTA 2196  
QY 2236 AAGGTGAGAGATGACATGAGTCTTATGGCTGCTTTCAGAAAGAAAGAAATG 2295  
Db 2197 AAGGTGAGAGATGACATGAGTCTTATGGCTGCTTTCAGAAAGAAAGAAATG 2256  
QY 2296 ATGGAAGAAATGCAACAGCAACCTGTGAGCCATAGCGTTTACAGAGTCCCATAC 2355  
Db 2257 ATGGAAGAAATGCAACAGCAACCTGTGAGCCATAGCGTTTACAGAGTCCCATAC 2316  
QY 2356 CAGTTCTGCAATGTGTGAGAGATTGGCTTTCAGAAAGATGACTGCACACTTGCAT 2415  
Db 2317 CAGTTCTGCAATGTGTGAGAGATTGGCTTTCAGAAAGATGACTGCACACTTGCAT 2376

RESULT 2  
AL568673/c 993 bp mRNA linear EST 16-FEB-2001  
LOCUS AL568673 LTL.FL002.Pl1 Homo sapiens cDNA clone CS0DE004C14 3 prime  
DEFINITION mRNA sequence.  
ACCESSION AL568673  
VERSION AL568673.1 GI:12923248  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

Location/Qualifiers  
1..993  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="LTL.FL002.Pl1"  
/lab\_host="DH10B"  
/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand  
cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-stranded cDNA was digested with Not I  
and cloned into the Not I and Eco RV sites of the  
pCMVSPORT 6 vector. Library was constructed by Life  
Technologies. Contact : Feng Liang Life Technologies, a  
division of Invitrogen 9800 Medical Center Drive Rockville  
Maryland 20850. USA Fax : (1) 301 610 8371 Email :  
filiang@lifetech.com URL :  
http://fulllength.invitrogen.com"

## BASE COUNT

290 a 215 c 190 g 289 t 9 others

## ORIGIN

Query Match 30.0%; Score 973; DB 9; Length 993;  
Best Local Similarity 98.8%; Pred. No. 1,7e-242;  
Matches 981; Conservative 8; Mismatches 3; Indels 1; Gaps 1;

QY 2206 CAGTTTGAATAATGTGTTTTCAGATTCCTAAGGAGGAGAGATAGACATAGAGTCTT 2265  
Db 993 CAGTTTGAATAATGTGTTTTCAGATTCCTAAGGAGGAGAGATAGACATAGAGTCTT 934  
QY 2266 ATGGCTGCTTTCAGAAAGAGAA- GAAATGATGGAAGAAAACTGCACAGCAACTGT 2324  
Db 993 ATGGCTGCTTTCAGAAAGAGAAAGAAATGATGGAAGAAAACTGCACAGCAACTGT 874  
QY 2325 GAGCATATGAGCTGTTTACGCAAGTCCCATACCACTTCTGCAATGTGTTATGCAAGTTGG 2384  
Db 873 GAGCATATGAGCTGTTTACGCAAGTCCCATACCACTTCTGCAATGTGTTATGCAAGTTGG 814  
QY 2385 CTTTCAAGAAATGTAAGTACGACACCTTGGATTCGCAAAATACGAGCTGATATCTTAC 2444  
Db 813 CTTTCAAGAAATGTAAGTACGACACCTTGGATTCGCAAAATACGAGCTGATATCTTAC 754  
QY 2445 CAAGAACTCAAAAAACCTGGCAGAGAAAGCCAGAAAAATCTGCTGCATTAAGCTGAT 2504  
Db 753 CAAGAACTCAAAAAATCCTGGCAGAGAAAGCCAGAAAAATCTGCTGCATTAAGCTGAT 694  
QY 2505 CTATGTGTTTGAAGCTGTAAGTACACAGGCTTCTTCTGCCAGAGGACATCCGTTAAATAT 2564  
Db 693 CTATGTGTTTGAAGCTGTAAGTACACAGGCTTCTTCTGCCAGAGGACATCCGTTAAATAT 634  
QY 2565 TGTTCCTCCACACAGTACCTGAGAGCTATGATGATGATGATGATGATGATGATGATGAT 2624  
Db 633 TGTTCCTCCACACAGTACCTGAGAGCTATGATGATGATGATGATGATGATGATGATGAT 574  
QY 2625 CTCCAGCCCTGAACCTTGTATTTTATAGTGCATGCGAGCTATACCTCAGTATTTGTG 2684  
Db 573 CTCCAGCCCTGAACCTTGTATTTTATAGTGCATGCGAGCTATACCTCAGTATTTGTG 514  
QY 2685 GACATGACCCAGAAATTTGTAACATGCAAGATTTACATCAGAGCAACATGAGACCTT 2744  
Db 513 GACATGACCCAGAAATTTGTAACATGCAAGATTTACATCAGAGCAACATGAGACCTT 454  
QY 2745 TGCAAGCATCTTGGAGAGGATTCGCAAGTGGAGCCCTGTTGATTAATCTCATCA 2804  
Db 453 TGCAAGCATCTTGGAGAGGATTCGCAAGTGGAGCCCTGTTGATTAATCTCATCA 394  
QY 2805 TTTTAACAGCTGTGATGGCTTACCTTGGTGAACCTAACCAATATATGACATGATGCG 2864  
Db 393 TTTTAACAGCTGTGATGGCTTACCTTGGTGAACCTAACCAATATATGACATGATGCG 334  
QY 2865 TCAAGAGTGGCTGTAATATATCCATGGTATGATGATGATGATGATGATGATGATGAT 2924  
Db 333 TCAAGAGTGGCTGTAATATATCCATGGTATGATGATGATGATGATGATGATGATGAT 274  
QY 2925 AGGATACCCACATCTAGCATCTTATGCTTATGCTGCTTATATGCTTGGGGTGG 2984  
Db 273 AGGATACCCACATCTAGCATCTTATGCTTATGCTGCTTATATGCTTGGGGTGG 214  
QY 2985 GTAGTAGATACCAATATAAACAACCTTACAGACCTTCTCTCTGAGTGTCTTTA 3044  
Db 213 GTAGTAGATACCAATATAAACAACCTTACAGACCTTCTCTCTGAGTGTCTTTA 154  
QY 3045 ATTCCTTACTAGAGGAGATTAATATTTGATGATTAATGAAGAAATTTTCTGTATAT 3104  
Db 153 ATTCCTTACTAGAGGAGATTAATATTTGATGATTAATGAAGAAATTTTCTGTATAT 94  
QY 3105 AACGAGGCTTTTATTTTCTTAATAATGATAGTATTAATAATGTTAGATTAACGAATG 3164  
Db 93 AACGAGGCTTTTATTTTCTTAATAATGATAGTATTAATAATGTTAGATTAACGAATG 34  
QY 3165 ATTTTATTTTCCAGACATATTTTAAAGTGC 3197  
Db 33 ATTTTATTTTCCAGACATATTTTAAATATGC 1

LOCUS	AL540976	965 bp	mrna	linear	EST 16-FEB-2001
DEFINITION	AL540976.LTI_FL002_P11 Homo sapiens cDNA clone CS0DE004YC14 5 prime				
ACCESSION	AL540976				
VERSION	AL540976.1				
KEYWORDS	GI:12871606				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 965)				
TITLE	Li, W.-B., Gruber, C., Jesse, J., and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001)				
FEATURES	Contact: Genoscope				
SOURCE	Genoscope - Centre National de Sequencage				
	BP 191 91006 EVRY cedex - France				
	Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
	Location/Qualifiers				
	1..965				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DE004YC14"				
	/clone_id="LTI_FL002_P11"				
	/lab_host="DH10B"				
	/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-colligod(r) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"				
BASE COUNT	280 a 194 c 243 g 247 t 1 others				
ORIGIN					
Query Match	28.4%: Score 922; DB 9; Length 965;				
Best Local Similarity	99.5%: Pred. No. 3.3e-229;				
Matches 955; Conservative	1; Mismatches 1; Indels 3; Gaps 3;				
QY	137	CAGAAGTGGGCCACCATATCTGGAACACTACAGTCTATGCTTTGGAAGCGCAAAAGGGAATA	196		
Db	8	CGGAAGTGGGCCACCATATCTGGAACACTACAGTCTATGCTTTGGAAGCGCAAAAGGGAATA	67		
QY	197	AACATTAAAGACTCCCCCGGGGACCTGGAGAGTGCATTTTCCATGGTGGCGGAGCAG	256		
Db	68	AACATTAAAGACTCCCCCGGGGACCTGGAGAGTGCATTTTCCATGGTGGCGGAGCAG	127		
QY	257	CAGCTTACATGAATAAATCAGGTAGAGTTACTCTGCTCACTCTGTTTTCAGAAAGCT	316		
Db	128	CAGCTTACATGAATAAATCAGGTAGAGTTACTCTGCTCACTCTGTTTTCAGAAAGCT	187		
QY	317	TTTGCTCAGATCTTTCCTCAGTGGAGAAAGGGGAATATACAGAAGATGTCTCCCTCAAGT	376		
Db	188	TTTGCTCAGATCTTTCCTCAGTGGAGAAAGGGGAATATACAGAAGATGTCTCCCTCAAGT	247		
QY	377	GCTCAGACACGTGTCCTTTGGAGAAACTATAGTTGGCAATTCCTCAATTAACCAATG	436		
Db	248	GCTCAGACACGTGTCCTTTGGAGAAACTATAGTTGGCAATTCCTCAATTAACCAATG	307		
QY	437	ACTTCAAAATTTTAAAAATATGAGCTCAGCTTGTGAAGTCCCTCCAGATAAGTTG	496		
Db	308	ACTTCAAAATTTTAAAAATATGAGCTCAGCTTGTGAAGTCCCTCCAGATAAGTTG	366		
QY	497	GCTGTATCTACCCGTGCTCTCCAGTTACAGAAAGCAACAGCAAAATCTCGCAAGTGT	556		
Db	367	GCTGTATCTACCCGTGCTCTCCAGTTACAGAAAGCAACAGCAAAATCTCGCAAGTGT	426		

Oy	557	TCACAAAATAGCTCAGTCCGTAAGGTATACAGTTCACGCTCGAAAGATGCCTCACACAC	616
Dd	427	TGAGAAAAATGCTACTCTTAGATTACAGTTATCATAGTCTGGAAAGATGACTCACACAC	486
Oy	617	ATGCTGTTGATGCTGCTGGTGGAATGCCACCATGAAGAATCTTCTGCATGGGGAGGCTTG	676
Dd	487	ATGCGTTGATGCTGCTGGTGGAATGCACGCCAATGAAGATCTTCTGCATGGGGAGGCTTG	546
Oy	677	CCCTGGCCCCGTGTAAAGCTGGTGGAATTTCAATTCACAAGAAGACCAACAGTTTGTG	736
Dd	547	CCCTGG-CCGTGGTAAAAGCTGGTGGAATTTCAATTCACAAGAAGACCAACAGTTTGTG	605
Oy	727	CCAGATATGTTAAAGTGCACAGCTGTTAGATACCTGTCACAGGAGCAGGAGCTTCCCT	796
Dd	606	CCAGATATGTTAAAGTGCACAGCTGTTAGATACCTGTCACAGGAGCAGGAGCTTCCCT	665
Oy	797	GCAAAACAGATCATCATCTGTTGGGCGCTGGTGATGGAATGGATTAACACAGGATGTA	856
Dd	666	GCAAAACAGATCATCATCTGTTGGGCGCTGGTGATGGAATGGATTAACACAGGATGTA	725
Oy	857	CTGGAAGCTCACAAGGCCATTTAGTATTTCTGAATTTGTCATCTATAAAATACAC	916
Dd	726	CTGGAAGCTCACAAGGCCATTTAGTATTTCTGAATTTGTCATCTATAAAATACAC	785
Oy	917	ACATTAAACAGTGCATATTCACACCTTGGAGCTCTGGAGTTTTCAGTCCCTCTGAAT	976
Dd	786	ACATTAAACAGTGCATATTCACACCTTGGAGCTCTGGAGTTTTCAGTCCCTCTGAAT	845
Oy	977	TGTGTAC-AAAGACTATTGTAGAGACTATCCGGGTAGATTGCAAGGAGAACCAATGATG	1035
Dd	846	TGTGTACAAAGACTATTGTAGAGACTATCCGGGTAGATTGCAAGGAGAACCAATGATG	905
Oy	1036	AGTATTTGAAAGAAATTCACCTGGTGTAGACAATGAGAGACCCCTGCTGCTTAA	1095
Dd	906	AGTATTTGAAAGAAATTCACCTGGTGTAGACAATGAGAGACCCCTGCTGCTTAA	965

RESULT 4  
Bg680130                      852 bp       mRNA       linear       EST 01-MAY-2001

LOCUS                      60262828Bf1 NCL\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4753388 5' ,  
DEFINITION                  mRNA sequence.

ACCESSION                   Bg680130  
KEYWORDS                   Bg680130.1 GI:13911527  
SOURCE                     EST.  
ORGANISM                   human.  
                             Homo sapiens  
                             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
                             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE                   1 (bases 1 to 852)  
AUTHORS                    NIH-MGC http://mgc.nci.nih.gov/  
TITLE                       National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL                     Unpublished (1999)  
COMMENT                    Contact: Robert Strausberg, Ph.D.  
                             Email: cgabs-remail.nih.gov  
                             Tissue Procurement: James Cleaver, M.D.  
                             CDNA Library Preparation: Life Technologies, Inc.  
                             CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
                             DNA Sequencing by: Incyte Genomics, Inc.  
                             Clone distribution: MGC clone distribution information can be  
                             found through the I.M.A.G.E. Consortium/LLNL at:  
                             http://image.llnl.gov  
                             Plate: LLAM10612 row: O column: 21  
High quality sequence stop: 792.  
Location/Qualifiers  
1. 852  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4753388"  
/clone\_1ib="NCL\_CGAP\_Skn4"  
/tissue\_type="squamous cell carcinoma"  
/lab\_host="DH10B (r1 phage-resistant)"

FEATURES  
source

/note="Organ: skin; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI-CGAP Library."  
 BASE COUNT 275 a 177 c 204 g 196 t  
 ORIGIN

Query Match 22.6%; Score 733.4; DB 12; Length 852;  
 Best Local Similarity 97.0%; Pred. No. 5.3e-180;  
 Matches 822; Conservative 0; Mismatches 16; Indels 9; Gaps 7;

QY 1786 ATGGGATTCAACGTGGAAGAGATGTATGAGGCCACGCGATGGATCCAAAGATCCYGAAT 1845  
 |||||||  
 Db 1 ATGGGATTCAACGTGGAAGAGATGTATGAGGCCACGCGATGGATCCAAAGATCCYGAAT 60  
 QY 1846 CTCGGAACCCACACATATGAGAAATATCATATCTGTACTCTGGGAGAAAGAACAT 1905  
 |||||||  
 Db 61 CTCGGAACCCACACATATGAGAAATATCATATCTGTACTCTGGGAGAAAGAACAT 120  
 QY 1906 GACATTTGTCTCAGCTTCAGAAACTTCAAGTGTCTCCATCAGACAGAAATTTACGCCCA 1965  
 |||||||  
 Db 121 GACATTTGTCTCAGCTTCAGAAACTTCAAGTGTCTCCATCAGACAGAAATTTACGCCCA 180  
 QY 1966 GGAAGCAGAGATTAGAGATTGAAGAGCCCGGGCTGACCTTATTGAGGTGTTATGAAC 2025  
 |||||||  
 Db 181 GGAAGCAGAGATTAGAGATTGAAGAGCCCGGGCTGACCTTATTGAGGTGTTATGAAC 240  
 QY 2026 ATTGAAGATATGCTTTGTTAAATACAGAGAGAAATGCGAAGAAAGAGAGAGCCCTT 2085  
 |||||||  
 Db 241 ATTGAAGATATGCTTTGTTAAATACAGAGAGAAATGCGAAGAAAGAGAGAGCCCTT 300  
 QY 2086 TGGCCCTGTTAGGAGACTGAGCTATTTCAGACAAACAAACCCAGACGAATGAAGAA 2145  
 |||||||  
 Db 301 TGGCCCTGTTAGGAGACTGAGCTATTTCAGACAAACAAACCCAGACGAATGAAGAA 360  
 QY 2146 AATATCATATTTCTGAATATGCTCTGCTCCAACTCAAGAGCTTCTAGATCAAAAAGAA 2205  
 |||||||  
 Db 361 AATATCATATTTCTGAATATGCTCTGCTCCAACTCAAGAGCTTCTAGATCAAAAAGAA 420  
 QY 2206 CAGTTTGAAGAA- AATGCTGTTGAGGTTCTTAA- AGGTGAGAAAGATAGACATGAGCTCC 2263  
 |||||||  
 Db 421 CAGTTTGAAGAAATGCTGTTGAGGTTCTTAA- AGGTGAGAAAGATAGACATGAGCTCC 480  
 QY 2264 TTATGGCTGCTTTCAGAAAGAAAGAAATGATGGAAGAAACTGCACAGCAACCTG 2323  
 |||||||  
 Db 481 TTATGGCTGCTTTCAGAAAGAAAGAAATGATGGAAGAAACTGCACAGCAACCTG 540  
 QY 2324 TGAGCCATAGGCTGTTTACAGCAAGTCCCATATCAGTTCGCAATGTGTTAGCAGATTG 2383  
 |||||||  
 Db 541 TGAGCCATAGGCTGTTTACAGCAAGTCCCATATCAGTTCGCAATGTGTTAGCAGATTG 600  
 QY 2384 GCTTTC- AAGAATGATGCGACACCTTGCGATCCAAATATGAGAGCTGATATACCTTC 2442  
 |||||||  
 Db 601 GCTTTCACAGATGATGATGCGACACCTTGCGATCCAAATATGAGAGCTGATATACCTTC 660  
 QY 2443 ACCAAGAACCCTCAAAACCTGCGCAGAG- AAGGCCAAGAAATCTCTGCT- -GCAGATTAAG 2499  
 |||||||  
 Db 661 ACCAAGAACCCTCAAAACCTGCGCAGAGCAAGGCCAAGAAATCTCTGCTTGCAGATTAAAG 720  
 QY 2500 CTGATCTATGTGTTT- GAGGCTGAAGTACTACAGGCTTCTTCTGCCAGGACATCCGTT 2558  
 |||||||  
 Db 721 CTGATCTATGTGTTT- GAGGCTGAAGTACTACAGGCTTCTTCTGCCAGGACATCCGTT 780  
 QY 2559 AATATTTGTCCCGACACAGTACGCTGAGCTATATA- - GATGTCATGACAGTGTGTT 2616  
 |||||||  
 Db 781 AATATTTGTCCCGACACAGTACGCTGAGCTATACAGATGAGGCAATCACCAGGTTGTT 840  
 QY 2617 GACAAATG 2623  
 |||||  
 Db 841 GACCAAG 847

RESULT 5

B0014467/c  
 LOCUS B0014467 724 bp mRNA linear EST 26-MAR-2002  
 DEFINITION U1-H-ED1-axf-k-07-0-UI.s1 NCI-CGAP-ED1 Homo sapiens cDNA clone  
 IMAGE:5833470 3', mRNA sequence.  
 ACCESSION B0014467  
 VERSION B0014467.1 GI:19739368  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 724)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Seq primer: M13 FORWARD  
 POLY-A=yes.  
 FEATURES  
 source  
 location/Qualifiers  
 1..724  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5833470"  
 /clone\_lib="NCI CGAP-ED1"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Organ: Left Public Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; NCI CGAP-ED1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma cell line C5. The library was constructed according to Bonaldo , Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAGAGCT.  
 TAG\_Lib=UI-H-ED1  
 TAG\_Tissue=chondrosarcoma  
 TAG\_SEQ=CGTCAAGGCT"

BASE COUNT 227 a 145 c 135 g 216 t 1 others  
 ORIGIN

Query Match 22.0%; Score 713.6; DB 14; Length 724;  
 Best Local Similarity 99.3%; Pred. No. 7.3e-175;  
 Matches 716; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2523 AGTACTCAGAGCTTCTCTGCGAGGACATCCGTTAAATATGTTCCCCACACCTGAG 2582  
 |||||||  
 Db 724 AGTACTCAGAGCTTCTCTGCGAGGACATCCGTTAAATATGTTCCCCACACCTGAG 665  
 QY 2583 TCCCTGAGACTAATAGATGTCATGACAGAGTGTGTACAAATGCTCCACCCGGAACCTT 2642  
 |||||||  
 Db 664 TCCCTGAGACTAATAGATGTCATGACAGAGTGTGTACAAATGCTCCACCCGGAACCTT 605  
 QY 2643 TGTATATTTTATGTCATGACAGGCTATACCTCAGATTTTGTGACATGCACCCAGAGATA 2702  
 |||||||  
 Db 604 TGTATATTTTATGTCATGACAGGCTATACCTCAGATTTTGTGACATGCACCCAGAGATA 545  
 QY 2703 TGTACAGTCACAGATTACTATCAGAGCAATGAGACCCTTTGACAGCATCTCTTGAG 2762  
 |||||||

Query Match	Best Local Similarity	21.8%	Score 705.8	DB 12	Length 742
Matches 721	Conservative	98.8%	Pred. No. 8e-173	Mismatches 8	Indels 1
				Gaps 1	
QY 2357	AGTTCTGCATGTGGTATGACAGAGTTGGCTTTCAAGAATGTACTGCACACCTTGCATC				2416
Db 1	AGTTCTGCATGTGGTATGACAGAGTTGGCTTTCAAGAATGTACTGCACACCTTGCATC				60
QY 2417	CAAAATACGGAGCTGGCATTACTTCACCAAGAACCCTAAAACCTGGCAGAGAGGCCA				2476
Db 61	CAAAATACGGAGCTGGCATTACTTCACCAAGAACCCTAAAACCTGGCAGAGAGGCCA				120
QY 2477	AGAAATCTGTGCTGTCACATTAAGCTGACTATGTGTTGAGGCTGAAGTACTACAGGCT				2536
Db 121	AGAAATCTGTGCTGTCACATTAAGCTGACTATGTGTTGAGGCTGAAGTACTACAGGCT				180
QY 2537	TCTTTGCGCAGGACATCCGTTAAATATTGTTCCCCACACACTGAGTCTTGGAGCTATAG				2596
Db 181	TCTTTGCGCA-GGACATCCGTTAAATATTGTTCCCCACACACTGAGTCTTGGAGCTATAG				239
QY 2597	ATGGTCATGACAGTGTGGTTGCACATGTCTCCAGGCCCTGAAACCTTTGTTATTTTAAAG				2656
Db 240	ATGGTCATGACAGTGTGGTTGCACATGTCTCCAGGCCCTGAAACCTTTGTTATTTTAAAG				299
QY 2657	GCATCAGGCTTATACCTCAGTATTTTGTGACATGACACCAGGAATATGTACAGTACAAAG				2716
Db 300	GCATCAGGCTTATACCTCAGTATTTTGTGACATGACACCAGGAATATGTACAGTACAAAG				359
QY 2717	ATTACTCATCAGGACCAATGAGACCCCTTGGCAGACATCTTGGAGGGGATTCGCAAGTG				2776
Db 360	ATTACTCATCAGGACCAATGAGACCCCTTGGCAGACATCTTGGAGGGGATTCGCAAGTG				419
QY 2777	GCAGCCCTGTGTAAATCTCTACATCATTTTAAACAGCTGTGATGCGCTTACCTTGGGTG				2836
Db 420	GCAGCCCTGTGTAAATCTCTACATCATTTTAAACAGCTGTGATGCGCTTACCTTGGGTG				479
QY 2837	AACTAACCAAAATTAAGACATCGATGGCTCAAGAGTGGCTGAAATATATCCATGGGTT				2896
Db 480	AACTAACCAAAATTAAGACATCGATGGCTCAAGAGTGGCTGAAATATATCCATGGGTT				539
QY 2897	ATCTGTATGAGCTGACTGGTTATTGAAAGGACTAGCCACATCTAGCATCTTAGTGCT				2956
Db 540	ATCTGTATGAGCTGACTGGTTATTGAAAGGACTAGCCACATCTAGCATCTTAGTGCT				599
QY 2957	TTATCTGCTTATGCTCTTGGGCTGGGGTGGTATGATACCAATGAAACCTTTCAGGA				3016
Db 600	TTATCTGCTTATGCTCTTGGGCTGGGGTGGTATGATACCAATGAAACCTTTCAGGA				659
QY 3017	CGTTCCCTCCTCTGTCAGTTGTTCTTTTATCTCCTTTACTAGAGAGATAAATATTTGGC				3076
Db 660	CGTTCCCTCCTCTGTCAGTTGTTCTTTTATCTCCTTTACTAGAGAGATAAATATTTGGC				719
QY 3077	ATATATATGAA 3086				
Db 720	ATATGAGAA 729				

RESULT 7	LOCUS	901 bp	mRNA	linear	EST 20-AUG-2002
BQ923651	BQ923651				

```

DEFINITION AGENCOURT_8854013 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6473545
ACCESSION BQ923651
VERSION BQ923651.1 GI:22338682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LAM14009 row: e column: 02
High quality sequence stop: 583.
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6473545"
/lssue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 266 a 181 c 234 g 220 t
ORIGIN
Query Match 20.3%; Score 659.6; DB 14; Length 901;
Best Local Similarity 87.0%; Pred. No. 9.8e-161;
Matches 840; Conservative 0; Mismatches 14; Indels 112; Gaps 6;
QY 155 TCTGGAAACTACAGCTATGCTTTGAACGCGCAAAAGGGAATTAACATTTAAGACTCCC 214
Db 7 TCTGGAACACTACAGCTATGCTTTGAACGCGCAAAAGGGAATTAACATTTAAGACTCCC 66
QY 215 CGGGAGCTGGAGATGAGCTTTCCATGCTGGCGGAGCGAGCAGCTTACATGAAAAT 274
Db 67 CGGGAGCTGGAGATGAGCTTTCCATGCTGGCGGAGCGAGCAGCTTACATGAAAAT 125
QY 275 CAGTAGATTACCTCGCTCAGCTCTGTTTCAAGAAAGTCTTGGCTCAGATCTTCTC 334
Db 126 ----- 125
QY 335 AGTGAGAAAGGGAAATACAGAAAGATGTCCTTACAGTGTCTCAGAGACTGGTCTC 394
Db 126 -----TCAGAGACTGGTCTC 141
QY 395 TTGGAGAAACTATATGTTGGCAAAATCCCATTAACCAATGACTTCAAAATTTTAAAA 454
Db 142 TTGGAGAAACTATATGTTGGCAAAATCCCATTAACCAATGACTTCAAAATTTTAAAA 201
QY 455 ATAATGAGCGTACGCTGTGTAAGTCCCTCCAGAAATTAAGTTGGCTGTATCTACCCGG 514
Db 202 ATAATGAGCGTACGCTGTGTAAGTCCCTCCAGAAATTAAGTTGGCTGTATCTACCCGG 261
QY 515 TCTCTCCAGTTTCAGGAAGGCAACAGCAAAATCTGCAAGTGTGCAAGAAAATGCTGACTC 574
Db 262 TCTCTCCAGTTTCAGGAAGGCAACAGCAAAATCTGCAAGTGTGCAAGAAAATGCTGACTC 321
QY 575 CTAGAGTAGAGTTATCACTGTGGAAGATGACCTCACCACATGCTGTTGATGCTGTGG 634
Db 322 CTAGAGTAGAGTTATCACTGTGGAAGATGACCTCACCACATGCTGTTGATGCTGTGG 381

```

```

QY 635 TGAATGAGCCCAATGAAAGATCTTCTGCATGGGGAGGCGCTGGCCCTGGTAAAG 694
Db 382 TGATGAGCAACATGAAAGATCTTCTGCATGGGGAGGCGCTGGCCCTGGTAAAG 441
QY 695 CTGCTGATTTTGAATCCAGAGAGCAACAGTTTGTTCAGATATGTAAGTGT 754
Db 442 CTGCTGATTTTGAATCCAGAGAGCAACAGTTTGTTCAGATATGTAAGTGT 501
QY 755 CAGCTGTGATATGCTGTACGGGAGAGGAGGAGGCTCCCTGCAACAGATCATCATG 814
Db 502 CAGCTGTGATATGCTGTACGGGAGAGGAGGAGGCTCCCTGCAACAGATCATCATG 561
QY 815 CTGTTGGGCGCTGGTGAATGGAATGGAATTAACAGGATGACTGGAAGCTGAGAGG 874
Db 562 CTGTTGGGCGCTGGTGAATGGAATGGAATTAACAGGATGACTGGAAGCTGAGAGG 621
QY 875 CCATTTGATATTTCTGAATATGATGATCTATATAAAATCTCAGATTAAGACAGTAA 934
Db 622 CCATTTGATATTTCTGAATATGATGATCTATATAAAATCTCAGATTAAGACAGTAA 680
QY 935 TTCCAGCTTGAAGCTCTGGGATTTTTCAGTTCCTCTGATTTTGTACAAAGCTATTG 994
Db 681 TTCCAGCTTGAAGCTCTGGGATTTTTCAGTTCCTCTGATTTTGTACAAAGCTATTG 740
QY 995 TAGAGACTAT-CCGGGTAGTTGTCAGAGGAGCAATGATGA-GTAATTTGAAGAAAT 1052
Db 741 TAGAGACTATCCGGGTAGTTGTCAGAGGAGCAATGATGATGATTTGAAGAAAT 800
QY 1053 TCACCTGGTGAGCAAT-GAGGAGCCCTACTGTTG--CTGCTTTAAGCTGCTTACAGAT 1108
Db 801 TCACCTGGTGAGCAATGAGGAGCCCTACTGTTGCTGCTTTAAGCTGCTTCAAAAT 860
QY 1109 TCATCC 1114
Db 861 TCCTCC 866

```

```

RESULT 8
LOCUS BG284006 791 bp mRNA linear EST 21-FEB-2001
DEFINITION 602407302F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4519248 5',
ACCESSION BG284006 mRNA sequence.
VERSION BG284006.1 GI:13034520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 791)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-remail.nih.gov
Tissue Procurement: DCTD/DPF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LAM10415 row: d column: 01
High quality sequence stop: 729.
Location/Qualifiers
1..791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4519248"
/lssue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
FEATURES
source

```



```

Db 548 AATATGTACAGTCACACAGATTACTCATCAGACCAANGAGACCTTTCACAGCATCCTT 489
QY 2759 GGAGGGGATTCGCAAGTGGCAGCCCTGTTGATTAATCTCATCATCATTTTAACAGCTGTT 2818
Db 488 GGAGGGGATTCGCAAGTGGCAGCCCTGTTGATTAATCTCATCATCATTTTAACAGCTGTT 429
QY 2819 ATGGCCTTACCTGGGTGAGTAACTAACCAAAATATGACCATGAGCTTCAAGAGTGGCTT 2878
Db 428 ATGGCCTTACCTGGGTGAGTAACTAACCAAAATATGACCATGAGCTTCAAGAGTGGCTT 369
QY 2879 GAATATATCCCATGGGTATATCTGATATGACGACGCGGTATATGAAGAGTACGACAT 2938
Db 368 GAATATATCCCATGGGTATATCTGATATGACGACGCGGTATATGAAGAGTACGACAT 309
QY 2939 ACTAGCATCTAGTACCTTATGCTGCTTATGCTGCTTATGCTGCTTATGCTGCTTATGCTGCTT 2998
Db 308 ACTAGCATCTAGTACCTTATGCTGCTTATGCTGCTTATGCTGCTTATGCTGCTTATGCTGCTT 249
QY 2999 AATGAACACTTTCAGAGACCTTCCCTCTTCGAGTGTCTTAAATCTCTTACTAG 3058
Db 248 AATGAACACTTTCAGAGACCTTCCCTCTTCGAGTGTCTTAAATCTCTTACTAG 189
QY 3059 AGGAGATTAATATTTTGCATATATGAAGAATTTTCTAGTATATTAACGACGCTTTT 3118
Db 188 AGGAGATTAATATTTTGCATATATGAAGAATTTTCTAGTATATTAACGACGCTTTT 129
QY 3119 ATTTCTTAATATGATAGATATATAAATGTTAGATTAACGAATGATTTAGATTTCC 3178
Db 128 ATTTCTTAATATGATAGATATATAAATGTTAGATTAACGAATGATTTAGATTTCC 69
QY 3179 AGAGATTAATTAAGTCTTAGCTATGAATAATTAATCATCTTGTCTGATTAATAAAA 3238
Db 68 AGAGATTAATTAAGTCTTAGCTATGAATAATTAATCATCTTGTCTGATTAATAAAA 9
QY 3239 AAAAA 3243
Db 8 AAAAA 4

```

```

RESULT 10
LOCUS B0028103/c 658 bp mRNA linear EST 27-MAR-2002
DEFINITION UI-H-CO0-ari-b-03-0-UI.s1 NCI.CGAP_Sub9 Homo sapiens cDNA clone
IMAGE:3106491.3, mRNA sequence.
ACCESSION B0028103
VERSION B0028103.1 GI:19763382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 658)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Human Gene Index
Supplement 1 (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLY-A-ies.

```

```

FEATURES
Source
1. 658
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3106491"
/clone_lib="NCI_CGAP_Sub9"
/tissue_type="mixed"

```

```

/dev_stage="mixed"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker. Site.1: EcoR I; Site.2: Not I; tissues:
choleonic mucosa with Crohns disease, Choleonic mucosa with
ulcerative colitis, Fetal thymus, Cervix, Cervical
adenosquamous carcinoma, Ligament cells, Prostate
carcinoma, Bladder carcinoma, Brain oligodendrocyte ;
NCI CGAP Sub9 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GCTC, AACG, GGGCC,
GGAAG, TACG, TAAGC, ATGG, AGACA, ATGAC. For additional
information, contact: Bento Soares, bento-soares@iowa.edu
TAG-LIB-UI-H-CO0
TAG-TISSUE=Prostate Carcinoma
TAG_SEQ=ATGG*
BASE COUNT 147 a 156 c 118 g 235 t 2 others
ORIGIN
Query Match 19.9% Score 645; DB 14; Length 658;
Best Local Similarity 98.9%; Pred. No. 5.6e-157;
Matches 648; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1639 ACAGATTGGAGATATATTAAGGCTTTCAGTCTGAATGCGCAAGAGTCCCAAGTGTG 1698
Db 658 ACAAGATTGGAGATATATTAAGGCTTTCAGTCTGAATGCGCAAGAGTCCCAAGTGTG 599
QY 1699 AGTTGAAACAATTCACAGTGTCCCAAGTCAACGAGAGGAGAAAGAAATGGGCTT 1758
Db 598 AGTTGAAACAATTCACAGTGTCCCAAGTCAACGAGAGGAGAAAGAAATGGGCTT 539
QY 1759 GAAGCTAATCTCTCCCATCATCTGATGAGATTCACGTAAGAGATATGAGAGCC 1818
Db 538 GAAGCTAATCTCTCCCATCATCTGATGAGATTCACGTAAGAGATATGAGAGCC 479
QY 1819 CACGATGATCCAAAGAAATCTGAGTCCAGAACCCACATCATGGAATATCAT 1878
Db 478 CACGATGATCCAAAGAAATCTGAGTCCAGAACCCACATCATGGAATATCAT 419
QY 1879 ATTCGTACCTTGGGAGAAAGACATGACATTTTGTCTCAGCTTCAGAAAACCTCAAGT 1938
Db 418 ATTCGTACCTTGGGAGAAAGACATGACATTTTGTCTCAGCTTCAGAAAACCTCAAGT 359
QY 1939 GTCTCCATCAGAAATATATGACCCGAGAGAGGATGAGATTTGAAGAGCCCGG 1998
Db 358 GTCTCCATCAGAAATATATGACCCGAGAGAGGATGAGATTTGAAGAGCCCGG 299
QY 1999 GCTGACCTCATTTGAGGTGTTATGACATTTGAAGATGCTTTGAAGTACAGAGGAA 2058
Db 298 GCTGACCTCATTTGAGGTGTTATGACATTTGAAGATGCTTTGAAGTACAGAGGAA 239
QY 2059 ATGGCAAGAAAAGAGAGCCCTTGGCGCTGCTTGAAGACAGTGAATTCAGCAA 2118
Db 238 ATGGCAAGAAAAGAGAGCCCTTGGCGCTGCTTGAAGACAGTGAATTCAGCAA 179
QY 2119 CAAAAAACCCCAAGAGCAATGAAAGAAATATGATTTGGAATGCTGTCGCTCA 2178
Db 178 CAAAAAACCCCAAGAGCAATGAAAGAAATATGATTTGGAATGCTGTCGCTCA 119
QY 2179 ACTCAAGAGCTTTCAGATCAAAAGACATTTGAAGATGCTTTCAGATTCAGAA 2238
Db 118 ACTCAAGAGCTTTCAGATCAAAAGACATTTGAAGATGCTTTCAGATTCAGAA 59
QY 2239 GTGGAAGAGATGACATGAGTCTTATGCTGCTTTCAGAAAGAGAGAGAAAA 2293
Db 58 GTGGAAGAGATGACATGAGTCTTATGCTGCTTTCAGAAAGAGAGAGAAAA 4

```

RESULT 11  
BM829571  
LOCUS 608 bp mRNA linear EST 06-MAR-2002  
DEFINITION K-EST0102596 S9SNU601 Homo sapiens cDNA clone S9SNU601-59-H09 5',  
mRNA sequence.  
ACCESSION BM829571  
VERSION BM829571.1 GI:19185980  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 608)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 59 row: H column: 09  
High quality sequence stop: 608.  
Location/Qualifiers  
1..608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S9SNU601-59-H09"  
/clone\_lib="S9SNU601"  
/sex="M"  
/tissue\_type="ascites"  
/cell\_type="Epithelial"  
/cell\_line="SNU-601"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pME18-Fl3; Site:1: XhoI:  
Site:2: XhoI: The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DraIII- digested pME18S-Fl3 vector. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

BASE COUNT 196 a 114 c 132 g 166 t  
ORIGIN

Query Match 18.6%; Score 604.8; DB 14; Length 608;  
Best Local Similarity 99.7%; Pred. No. 1.7e-146;  
Matches 606; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 981 TACAAGACTATGTGAGACGCTCCGGGTAGTTGCAAGGAGCAATGATGAGTAA 1040  
|||||  
Db 1 TACAAGACTATGTGAGACGCTCCGGGTAGTTGCAAGGAGCAATGATGAGTAA 60  
|||||  
QY 1041 TTGGAAGAAATTCACCTGGTGAAGCAATGAGACCCCTACTGTGTGCTTAAAGCTGC 1100  
|||||  
Db 61 TTGGAAGAAATTCACCTGGTGAAGCAATGAGACCCCTACTGTGTGCTTAAAGCTGC 120  
|||||  
QY 1101 TTGAGATTCATCTTGAGGAGAGTGAGCTGGGACAGAAACACCCCTTTCTTCAATGC 1160  
|||||  
Db 121 TTGAGATTCATCTTGAGGAGAGTGAGCTGGGACAGAAACACCCCTTTCTTCAATGC 180  
|||||

QY 1161 AATGTCGTGACCAACCTGACCCCTCCAGATTGTCCAGGGCCACATTTGAATGGCAGACGC 1220  
|||||  
Db 181 AATGTCGTGACCAACCTGACCCCTCCAGATTGTCCAGGGCCACATTTGAATGGCAGACGC 240  
|||||  
QY 1221 AGATGTAATTTGTAATTTCTGTAAACCCACATGATATTACAGTTGGACCTGTGGCAAGTC 1280  
|||||  
Db 241 AGATGTAATTTGTAATTTCTGTAAACCCACATGATATTACAGTTGGACCTGTGGCAAGTC 300  
|||||  
QY 1281 AATTCACAAACAGCAGGAGTTGAATGAATCGAATTTCTGGCACAAGGCTAAACA 1340  
|||||  
Db 301 AATTCACAAACAGCAGGAGTTGAATGAATCGAATTTCTGGCACAAGGCTAAACA 360  
|||||  
QY 1341 GTTTCACAGCTCCCACTGCTGCTACGTCACAAAAGATTAACTTTGTAATATAT 1400  
|||||  
Db 361 GTTTCACAGCTCCCACTGCTGCTACGTCACAAAAGATTAACTTTGTAATATAT 420  
|||||  
QY 1401 ATACCATGTACTGTGCTGATTCAGAAATTTCTTAACCTCGATATTAACATGCAATGAA 1460  
|||||  
Db 421 ATACCATGTACTGTGCTGATTCAGAAATTTCTTAACCTCGATATTAACATGCAATGAA 480  
|||||  
QY 1461 GGAGCTTTGGAAAATGCAATGAGCAAAATATTAATTCATTTCTTCTGCTGCTGG 1520  
|||||  
Db 481 GGAGCTTTGGAAAATGCAATGAGCAAAATATTAATTCATTTCTTCTGCTGCTGG 540  
|||||  
QY 1521 GACTGGAACATGGAATTAAGAGGAACAGCAGCAGAGATTGTTGATGAACTTT 1580  
|||||  
Db 541 GACTGGAACATGGAATTAAGAGGAACAGCAGCAGAGATTGTTGATGAACTTT 600  
|||||  
QY 1581 AACATTG 1588  
|||||  
Db 601 AACATTG 608  
|||||

RESULT 12  
B1857151 705 bp mRNA linear EST 10-OCT-2001  
LOCUS 603383747F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:5392715 5',  
DEFINITION mRNA sequence.  
ACCESSION B1857151  
VERSION B1857151.1 GI:15997898  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 705)  
NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-rc@mail.nih.gov  
Tissue Procurement: DCTD/DMR  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM12000 row: n column: 12  
High quality sequence stop: 691.  
Location/Qualifiers  
1..705  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5392715"  
/clone\_lib="NIH\_MGC\_87"  
/tissue\_type="mammary adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: breast; Vector: pCMV-Sport6; Site:1: NotI:  
Site:2: SalI; Cloned unidirectional; oligo-dT primed.  
Average insert size 1.383 kb. Library enriched for  
full-length clones and constructed by Life Technologies."

BASE COUNT 226 a 121 c 151 g 207 t  
 ORIGIN

Note: this is a NIH-MGC Library."

Query Match 18.5% Score 600.4; DB 13; Length 705;  
 Best Local Similarity 98.4%; Pred. No. 2.5e-145;  
 Matches 659; Conservative 0; Mismatches 6; Indels 5; Gaps 5;

QY 2576 CACGAGCCGCGAGCTTACATGCTCATGACAGTGGTGGACAAATCTCCAGCCCTG 2635  
 DB 1 CACGAGCCGCGAGCTTACATGCTCATGACAGTGGTGGACAAATCTCCAGCCCTG 60  
 QY 2636 AACCTTGTATTTTGTGTCATGACGAGCTATACCTCAGTATTTGGACATGACCC 2695  
 DB 61 AAAC- TTGTTATTTTGTGTCATGACGAGCTATACCTCAGTATTTGGACATGACCC 118  
 QY 2696 AGGAATATGTACAGTACACAGATTACTCATGACGACCAATGAGACCCCTTGCACAGATC 2755  
 DB 119 AGGAATATGTACAGTACACAGATTACTCATGACGACCAATGAGACCCCTTGCACAGATC 178  
 QY 2756 CTGGAGGAGGATGGCAAGTGGCAGCCCTGTGATTTATCTCTACATCTTTTAACAGCT 2815  
 DB 179 CTGGAGGAGGATGGCAAGTGGCAGCCCTGTGATTTATCTCTACATCTTTTAACAGCT 238  
 QY 2816 GGTATGGCTTACCTTGGGTAACCTAACCAATATGACATGCGATGGCTCAAGAGTGG 2875  
 DB 239 GGTATGGCTTACCTTGGGTAACCTAACCAATATGACATGCGATGGCTCAAGAGTGG 298  
 QY 2876 CTGGAATATATCCATGCGGTTATCTGTATGAGTACTGAGTGGTATTTGAAGACTAGCCA 2935  
 DB 299 CTGGAATATATCCATGCGGTTATCTGTATGAGTACTGAGTGGTATTTGAAGACTAGCCA 358  
 QY 2936 CATCTACATCTTACGCTTATCTCTTATGCTTATGCTTGGGCTGGGCTAGTACAGTA 2995  
 DB 359 CATCTACATCTTACGCTTATCTCTTATGCTTATGCTTGGGCTGGGCTAGTACAGTA 418  
 QY 2996 CCAATATGAACACTTTCAGGACCTTCTCTTCTGACGTTGCTTAAATCTCCTTAC 3055  
 DB 419 CCAATATGAACACTTTCAGGACCTTCTCTTCTGACGTTGCTTAAATCTCCTTAC 478  
 QY 3056 TAGAGGAGATTAATTTTTCATATATGAGAAATTTTTCATATATTAACGAGGCT 3115  
 DB 479 TAGAGGAGATTAATTTTTCATATATGAGAAATTTTTCATATATTAACGAGGCT 537  
 QY 3116 TTTATTTCTTAAATGATGATATAT - AAAATGTTAGGATTAACGATGATTTAGATT 3174  
 DB 538 TTTATTTCTTAAATGATGATATTAATAAATGTTAGGATTAACGATGATTTAGATT 597  
 QY 3175 TTCCAGAAATATTTAAAGTCTTAAAGTATGAAAT - AAATCATCTTTGCTGATTAA 3233  
 DB 598 TTCCAGAAATATTTAAAGTCTTAAAGTATGAAATTAATTAATCATCTTTGCTGAAAAA 657  
 QY 3234 AAAAAAAAAA 3243  
 DB 658 AAAAAAAAAA 667

RESULT 13  
 AM851285/c 601 bp mRNA linear EST 19-MAY-2000  
 LOCUS AM851285  
 DEFINITION IL2-CT0220-170200-067-H02 CT0220 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION AM851285  
 VERSION AM851285.1 GI:7946802  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 601)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

TITLE 'M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 JOURNAL Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 COMMENT 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL2-CT0220-170  
 200-067-H02&t3=2000-02-17&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 600.

## FEATURES

## source

Location/Qualifiers  
 1..601  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="CT0220"  
 /dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 191 a 131 c 123 g 156 t  
 ORIGIN

Query Match 18.1%; Score 586.8; DB 10; Length 601;  
 Best Local Similarity 99.5%; Pred. No. 8.4e-142;  
 Matches 599; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 2506 TATGTGTTGAGGCTGAAGTACTACAGGCTTCTTCCAGGAGACATCCGTTAAATTT 2565  
 DB 601 TATGTGTTGAGGCTGAAGTACTACAGGCTTCTTCCAGGAGACATCCGTTAAATTT 542  
 QY 2566 GTTCCCCACACATGAGTCCGAGGCTATAGATGCTATGACAGTGGTGTGAACAATGTC 2625  
 DB 541 GTT-CCCCACACATGAGTCCGAGGCTATAGATGCTATGACAGTGGTGTGAACAATGTC 483  
 QY 2626 TCAGACCCCTGAACCTTGTATTTTGTGTCATGACGAGCTATACCTCAGTATTTGTGG 2685  
 DB 482 TCAGACCCCTGAACCTTGTATTTTGTGTCATGACGAGCTATACCTCAGTATTTGTGG 423  
 QY 2686 ACATGACCCAGGAATATGTACAGTACAGATTACTCATGACGACCAATGAGACCCCTT 2745  
 DB 422 ACATGACCCAGGAATATGTACAGTACAGATTACTCATGACGACCAATGAGACCCCTT 363  
 QY 2746 GCACGACCTCTGGAGGAGGATTCGACAGTGGAGCCCTGTTGATATCTCTCATCAT 2805  
 DB 362 GCACGACCTCTGGAGGAGGATTCGACAGTGGAGCCCTGTTGATATCTCTCATCAT 303  
 QY 2806 TTTAAGACCTGATGAGCTTACCTTGGGTGAAGTAACCAATATATGACATGATGCT 2865  
 DB 302 TTTAAGACCTGATGAGCTTACCTTGGGTGAAGTAACCAATATATGACATGATGCT 243  
 QY 2866 CAAGAGTGGCTTGAATATATCCATGAGGTTATCTGTATGACATGAGTGGCTATTGAAA 2925  
 DB 242 CAAGAGTGGCTTGAATATATCCATGAGGTTATCTGTATGACATGAGTGGCTATTGAAA 183  
 QY 2926 GGACTAGCACAATATACATCTTACGCTTATCTGCTTTATGCTTGGGCTGGGG 2985  
 DB 182 GGACTAGCACAATATACATCTTACGCTTATCTGCTTTATGCTTGGGCTGGGG 123  
 QY 2986 TAGTAGATACCAATGAACAACCTTCAGACCTTCCTCTGACAGTGTCTTTAA 3045

|||||  
Db 122 TAGGTAGATACCAATGAACACTTTCAGACCTTCCTCTCTGCTGCTTTTAA 63  
OY 3046 TCTCCTTTACTAGAGAGATTAATTTTGCATATATATAGAGAAATTTTCTAGTATA 3105  
Db 62 TCTCCTTTACTAGAGAGATTAATTTTGCATATATATAGAGAAATTTTCTAGTATA 3  
OY 3106 AC 3107  
Db 2 AC 1  
RESULT 14  
LOCUS B1836699 732 bp mRNA linear EST 04-OCT-2001  
DEFINITION 603089634F1 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE:5228481 5',  
mRNA sequence.  
ACCESSION B1836699  
VERSION B1836699.1 GI:15948249  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 732)  
AUTHORS NIH-MGC http://mgi.mgi.nhl.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Clone Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM11575 row: C column: 10  
High quality sequence stop: 732.  
Location/Qualifiers  
1. 732  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5228481"  
/clone\_lib="NIH\_MGC\_120"  
/lab\_host="DH10B"  
/note="Organ: pooled pancreas and spleen; Vector:  
pCMV-SPORT6; Site\_1: NOTI; Site\_2: EcoRV (destroyed); RNA  
source anonymous pool of spleen and pancreas from 28 yo  
male. Library is oligo-dT primed and directionally cloned  
(EcoRV site is destroyed upon cloning). Average insert  
size 1.5 kb, insert size range 1-2.5 kb. Library is  
normalized and enriched for full-length clones and was  
constructed by C. Gruber (Invitrogen). Research Genetics  
tracking code 023. Note: this is a NIH-MGC Library."  
BASE COUNT 213 a 150 c 187 g 182 t  
ORIGIN  
Query Match 18.1%; Score 586.2; DB 13; Length 732;  
Best Local Similarity 86.7%; Pred. No. 1.3e-141;  
Matches 726; Conservative 0; Mismatches 3; Indels 108; Gaps 3;  
OY 137 CAGAGTGGGCGACCATATCTGGAACACTACAGTCTATGCTTTGAAGCGCAAAAGGAAATA 196  
Db 1 CGGAAGTGGGCGACCATATCTGGAACACTACAGTCTATGCTTTGAAGCGCAAAAGGAAATA 60  
OY 197 AACATTTAAAGACTCCCCGGGAGCTGGAGATGAGCTTTTCATGCTGGCCGGAGACAG 256  
Db 61 AACATTTAAAGACTCCCCGGGAGCTGGAGATGAGCTTTTCATGCTGGCCGGAGACAG 120  
OY 257 CAGCTTACATGATAAATAGGTAGATTAAGTCTGCTGCTACTCTTGTTCAGAAAGTCT 316  
Db 121 CAGCTTACATGATAAATAGGTAGATTAAGTCTGCTGCTACTCTTGTTCAGAAAGTCT 137

OY 317 TTGCTCAGATCTTCTCCTCAGTGGAGAAAGGGAATACAGAGAATGTCTCCCTACAGT 376  
Db 138 ----- 137  
OY 377 GCTCAGAGACTGTGCTCTTGGAGAAACTATAGTTGGCAATTCOCATTAACCAATG 436  
Db 138 --TCAGAGACTGTGCTCTTGGAGAAACTATAGTTGGCAATTCOCATTAACCAATG 195  
OY 437 ACTTCAA-AATTTTAAATAATAGAGCTCAGCTGTGTGAAGTCTCCAGATAAGTT 495  
Db 196 ACTTCAACAATTTTAAATAATAGAGCTCAGCTGTGTGAAGTCTCCAGATAAGTT 255  
OY 496 GCGTGTATCTTACCTGCTGTCTCCAGTTCAAGAAAGCAACAGCAATCTCGCAAGT 555  
Db 256 GCGTGTATCTTACCTGCTGTCTCCAGTTCAAGAAAGCAACAGCAATCTCGCAAGT 315  
OY 556 TTCCA--GAAATAATGCTGATCCTAGATAGATTTATCACTGGAAGAAGTGAACCTCACA 613  
Db 316 TTCAATGAAATAATGCTGATCCTAGATAGATTTATCACTGGAAGAAGTGAACCTCACA 375  
OY 614 CACATGCTGTGATGCTGTGTGAATGCAGCAATGAAATCTTGTGATGGGGAGGCC 673  
Db 376 CACATGCTGTGATGCTGTGTGAATGCAGCAATGAAATCTTGTGATGGGGAGGCC 435  
OY 674 TGGCCCTGGCCCTGTGTAAGAGCTGTGATTTGAAATCCAGAAGAGACCAACAGTTTG 733  
Db 436 TGGCCCTGGCCCTGTGTAAGAGCTGTGATTTGAAATCCAGAAGAGACCAACAGTTTG 495  
OY 734 TTGCCAGATATGTGTAAGAGCTGTGATTTGAAATCCAGAAGAGACCAACAGTTTG 793  
Db 496 TTGCCAGATATGTGTAAGAGCTGTGATTTGAAATCCAGAAGAGACCAACAGTTTG 555  
OY 794 CCGCAACAGATATGATGCTGTGATTTGAAATCCAGAAGAGACCAACAGTTTG 853  
Db 556 CCGCAACAGATATGATGCTGTGATTTGAAATCCAGAAGAGACCAACAGTTTG 615  
OY 854 TTACTGGAAGCTGCGAGAGGCGCATTTGATTTCTGATTTGATTTGATTTGATTTGATTT 913  
Db 616 GTACTGGAAGCTGCGAGAGGCGCATTTGATTTCTGATTTGATTTGATTTGATTTGATTT 675  
OY 914 CTCACATTAAGACAGTACGATTTCCAGCTTGTGAGCTCTGGGATTTTTCAGTTCCTTC 970  
Db 676 CTCACATTAAGACAGTACGATTTCCAGCTTGTGAGCTCTGGGATTTTTCAGTTCCTTC 732  
RESULT 15  
LOCUS AWB51000 610 bp mRNA linear EST 19-MAY-2000  
DEFINITION IL3-CT0220-150200-069-E07 CT0220 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AWB51000  
VERSION AWB51000.1 GI:7946517  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 610)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verilovski-Almeida,S., Britones,M.R.,  
Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,L.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-il3-CT0220-150>)  
200-069-E07&L3=2000-02-15&t4=1)

Seq primer: puc 18 forward  
High quality sequence stop: 545.

## FEATURES

## Source

Location/Qualifiers  
1..610

/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone\_lib="CF0220"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 208 a 114 c 147 g 140 t 1 others  
ORIGIN

Query Match 18.0%; Score 584.2; DB 10; Length 610;  
Best Local Similarity 98.2%; Pred. No. 4e-141;

Matches 601; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

```
Oy 1776 CATCAATCTGATGGATTCACAGTGAAGAGATGTATGAGGCCACGCGATGATCCAAAG 1835
      |||||||
Db 1 CATCAATCTAATGGATTCAAGCTGGAAGAGATGTATGAGGCCACGCGATGATCCAAAG 60
Oy 1836 AATCTGAGTCTCCGAACCCACACATCATTTGAGATAATCATFATTTCTGTACCTTGGAG 1895
      |||||||
Db 61 AATCTGAGTCTCCGAACCCACACATCATTTGAGATAATCATFATTTCTGTACCTTGGAG 120
Oy 1896 AAAGGACATGATCTTTGTTCTCAGCTCAGAAAACCTCAAGTCTCCATCAGCAAGAAAT 1955
      |||||||
Db 121 AAAGGACATGATCTTTGTTCTCAGCTCAGAAAACCTCAAGTCTCCATCAGCAAGAAAT 180
Oy 1956 TATCAGCCAGGAGAGACAGATTGAGATTGAAGAGCCCGGCTGACCTCATTTGAGGT 2015
      |||||||
Db 181 TATCAGCCAGGAGAGACAGATTGAGATTGAAGAGCCCGGCTGACCTCATTTGAGGT 240
Oy 2016 GGTATGACATTTGAATATGCTTTGTAAGTACAGAGGAAATGCAAGGAAAAAGGA 2075
      |||||||
Db 241 GGTATGACATTTGAATATGCTTTGTAAGTACAGAGGAAATGCAAGGAAAAAGGA 300
Oy 2076 GCGAGGCGTTGGGCTGTTGGACAGTGCATTTGACGCAACAAAAAACCCAGACGA 2135
      |||||||
Db 301 GCGAGGCGTTGGGCTGTTGGACAGTGCATTTGACGCAACAAAAAACCCAGACGA 360
Oy 2136 AATGAAGAAAAATATCATATTTCTGAATATGCTGCTGCTCCACTCAAGAGCTTCTAGA 2195
      |||||||
Db 361 AATGAAGAAAAATATCATATTTCTGAATATGCTGCTGCTCCACTCAAGAGCTTCTAGA 420
Oy 2196 TCAAAAGAAACAGTTGAAAAATGTGTTGAGGTTCTAAAGGTGAGAGATAGACAA 2255
      |||||||
Db 421 TCAAAAGAAACAGTTGAAAAATGTGTTGAGGTTCTAAAGGTGAGAGATAGACAA 480
Oy 2256 TGAGGTCTTATGGCTGCTTCAAGAAAGAAAGAAATGATGSAAGAAAACTGCACAG 2315
      |||||||
Db 481 TGAGGTCTTATGGCTGCTTCAAGAAAGAAAGAAATGATGSAAGAAAACTGCACAG 538
Oy 2316 GCACCTGTGAGCCATAGCTGTTTACGCAAGTCCATACAGTCTGCAATGTGTATG 2375
      |||||||
Db 539 GCACCTGTGAGCCATAGCTGTTTACGCAAGTCCATACAGTCTGCAATGTGTATG 598
Oy 2376 CAGAGTTGGCTT 2387
      |||||||
Db 599 CAGAGTTGGCTT 610
```

Search completed: February 7, 2003, 05:35:47  
Job time : 2907 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2003, 23:46:53 ; Search time 5469 Seconds

(without alignments)  
17257.341 Million cell updates/sec

Title: US-09-830-762-1

Perfect score: 3243  
Sequence: 1 ggcctcgtgttcctcctgggtg.....gtctgattcaaaaaaaaaa 3243

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3241.4	100.0	3243	9 AF307338	AF307338 Homo sapi
2	3021.4	93.2	3138	9 AF307339	AF307339 Homo sapi
3	2842.6	87.7	3016	6 AX399212	AX399212 Sequence
4	2543.8	78.4	2598	6 AX274946	AX274946 Sequence
5	1440.4	44.4	1442	9 HSM803023	AL713679 Homo sapi
6	1371.8	42.3	1388	9 HSM804240	AL832929 Homo sapi
7	1362.4	42.0	3012	10 BC003281	BC003281 Mus muscu
8	844.4	26.0	167164	2 AC026506	AC026506 Homo sapi
9	844.4	26.0	174257	9 AC092908	AC092908 Homo sapi
10	821	25.3	167164	2 AC026506	AC026506 Homo sapi
11	622.8	19.2	822	6 AX188088	AX188088 Sequence
12	489.4	15.1	666	6 AX188241	AX188241 Sequence
13	463.2	14.3	516	6 AX185819	AX185819 Sequence
14	449.8	13.9	481	6 AX187250	AX187250 Sequence
15	414.8	12.8	457	6 AX184391	AX184391 Sequence
16	299	9.2	299	6 AX396731	AX396731 Sequence
17	284	8.8	284	6 AX332099	AX332099 Sequence
18	280.4	8.6	524	6 AX186264	AX186264 Sequence
19	246.8	7.6	260	6 AX330378	AX330378 Sequence
20	246.8	7.6	260	6 AX335102	AX335102 Sequence
21	236.6	7.3	124585	2 AC095176	AC095176 Rattus no
22	209.8	6.5	387	6 AX187472	AX187472 Sequence
23	202.8	6.3	214	6 AX302704	AX302704 Sequence
24	181.2	5.6	183	6 AX341343	AX341343 Sequence
25	140.4	4.3	5718	9 HSM804911	AL833598 Homo sapi
26	140.4	4.3	2223	9 AF484416	AF484416 Homo sapi
27	127.8	3.9	261	6 AX184601	AX184601 Sequence
28	122.8	3.8	2589	9 AK026003	AK026003 Homo sapi
29	122.8	3.8	2713	9 AK098816	AK098816 Homo sapi
30	122.8	3.8	5231	9 AB033094	AB033094 Homo sapi
31	122.8	3.8	149307	2 AC010858	AC010858 Homo sapi
32	122.8	3.8	203092	2 AC048348	AC048348 Homo sapi
33	120.6	3.7	66539	2 AC112130	AC112130 Homo sapi
34	120.6	3.7	66539	2 AC095519	AC095519 Rattus no
35	117.6	3.6	124585	2 AC095176	AC095176 Rattus no
36	82.6	2.5	10265	1 AE001727	AE001727 Thermotog
37	79	2.4	277050	1 AP000989	AP000989 Sulfolobu
38	75.6	2.3	11109	1 AE009804	AE009804 Pyrobacul
39	68.4	2.1	66539	2 AC095519	AC095519 Rattus no
40	68.2	2.1	96208	1 AL596174	AL596174 Listeria
41	68.2	2.1	311208	6 AX417049	AX417049 Sequence
42	68.2	2.1	349980	6 AX417036	AX417036 Sequence
43	65.8	2.0	3194	6 AX416716	AX416716 Sequence
44	65.4	2.0	7218	6 T66494	T66494 Sequence 14
45	62.6	1.9	225528	1 AL591984	AL591984 Listeria

#### ALIGNMENTS

RESULT 1  
AF307338  
LOCUS AF307338 3243 bp mRNA linear PRI 13-FEB-2001  
DEFINITION Homo sapiens B aggressive lymphoma long isoform (BAL) mRNA,  
complete cds.  
ACCESSION AF307338  
VERSION AF307338.1 GI:12751138  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1 (bases 1 to 3243)  
Aguilar,R.C., Yakushiji,Y., Kharbanda,S., Salgia,R., Fletcher,J.A.  
and Shipp,M.A.



QY 1501 ATTTCCTTCCCTGCGCTGGAGCTGAGAAACATGAAATTAAGAGAGAAACAGCAGCAGAG 1560  
 DB 1501 ATTTCCTTCCCTGCGCTGGAGCTGAGAAACATGAAATTAAGAGAGAAACAGCAGCAGAG 1560  
 QY 1561 ATTTTGTGTTGATGAAGTTTAAACATTTGGCAAGACCATGTAAGAACACCAAGTTAACTGTA 1620  
 DB 1561 ATTTTGTGTTGATGAAGTTTAAACATTTGGCAAGACCATGTAAGAACACCAAGTTAACTGTA 1620  
 QY 1621 AAATTTGATGATCTTCCCAAGATTTGGAGATATATTAAGGCTTTGAGTTCTTAATGGA 1680  
 DB 1621 AAATTTGATGATCTTCCCAAGATTTGGAGATATATTAAGGCTTTGAGTTCTTAATGGA 1680  
 QY 1681 AAGAGCTCCAGATGCTGAGTTTGAACAATTAAGCTGTCCCAATCAACAGAGAGAG 1740  
 DB 1681 AAGAGCTCCAGATGCTGAGTTTGAACAATTAAGCTGTCCCAATCAACAGAGAGAG 1740  
 QY 1741 AAAAGAGAAATGGCTTGAAGCTAGATCTCTGCAATCTGATGAGGATTTCAACGTG 1800  
 DB 1741 AAAAGAGAAATGGCTTGAAGCTAGATCTCTGCAATCTGATGAGGATTTCAACGTG 1800  
 QY 1801 GAAGAGATGATGAGGCGCCAGCATGATCCAAAGATCTGAGTCTCCAGAACCCAGC 1860  
 DB 1801 GAAGAGATGATGAGGCGCCAGCATGATCCAAAGATCTGAGTCTCCAGAACCCAGC 1860  
 QY 1861 ATCATTTGAGAAATATCATATTTCTGTACCTTGGAGAGAAAGAACATGACATTTTGTCTAG 1920  
 DB 1861 ATCATTTGAGAAATATCATATTTCTGTACCTTGGAGAGAAAGAACATGACATTTTGTCTAG 1920  
 QY 1921 CTTCAGAAAACCTTCAAGTGTCTCCATCACAGAAATTTATCAGCCCGAGAGACAGATTA 1980  
 DB 1921 CTTCAGAAAACCTTCAAGTGTCTCCATCACAGAAATTTATCAGCCCGAGAGACAGATTA 1980  
 QY 1981 GAGATTTGAAGAGCGCGGCTGACCTCATGAGGTGTTATGAATGAATGAATGAT 2040  
 DB 1981 GAGATTTGAAGAGCGCGGCTGACCTCATGAGGTGTTATGAATGAATGAATGAT 2040  
 QY 2041 TGTAAAGTACAGAGAGAAATGGCAAGAGAGAGCGAGCTTTGGCGCTGCTAGAGA 2100  
 DB 2041 TGTAAAGTACAGAGAGAAATGGCAAGAGAGAGCGAGCTTTGGCGCTGCTAGAGA 2100  
 QY 2101 CAGTGGACTATTCAGACACAAAAAACCCAGACGAATGAAGAAATATCATATTTCTG 2160  
 DB 2101 CAGTGGACTATTCAGACACAAAAAACCCAGACGAATGAAGAAATATCATATTTCTG 2160  
 QY 2161 AATGTGCTGCTGCTCCAACTCAAGAGCTTGAATCAAAAGAAACAGTTTGAAGAAATGT 2220  
 DB 2161 AATGTGCTGCTGCTCCAACTCAAGAGCTTGAATCAAAAGAAACAGTTTGAAGAAATGT 2220  
 QY 2221 GGTTCGACAGTCTTAAAGGTGAGAGAAATGAGCAATGAGCTTATGCTGCTTCA 2280  
 DB 2221 GGTTCGACAGTCTTAAAGGTGAGAGAAATGAGCAATGAGCTTATGCTGCTTCA 2280  
 QY 2281 AGAAGAGAGAAATATGATGAGAGAAACTGCAAGGCAACCTGTGAGCCATAGGCTGTT 2340  
 DB 2281 AGAAGAGAGAAATATGATGAGAGAAACTGCAAGGCAACCTGTGAGCCATAGGCTGTT 2340  
 QY 2341 CAGCAAGTCCCTACCACTGCTGCAATGTGATGCAAGGTTGGCTTTAAAGAAATGTC 2400  
 DB 2341 CAGCAAGTCCCTACCACTGCTGCAATGTGATGCAAGGTTGGCTTTAAAGAAATGTC 2400  
 QY 2401 TCGACACCTTGGATCCAAATATGAGAGCTGATATCTTCAACCAAGAACCTCAAAAC 2460  
 DB 2401 TCGACACCTTGGATCCAAATATGAGAGCTGATATCTTCAACCAAGAACCTCAAAAC 2460  
 QY 2461 CTGGCAGAGAGAGCCAGAAATCTCTGCTGCAAGATAGCTGATGTTGTTGAGGCT 2520  
 DB 2461 CTGGCAGAGAGAGCCAGAAATCTCTGCTGCAAGATAGCTGATGTTGTTGAGGCT 2520  
 QY 2521 GAAGTACTACAGAGCTTCTTTCGCAAGGACATCCGTTAAATATGTTCGCCACACAG 2580  
 DB 2521 GAAGTACTACAGAGCTTCTTTCGCAAGGACATCCGTTAAATATGTTCGCCACACAG 2580  
 QY 2581 AGTCTGAGAGCTATAGATGCTATGACAGTGTGTTGACAAATGTCTCAGCCCTGAAC 2640

DB 2581 AGTCTGAGAGCTATAGATGCTATGACAGTGTGTTGACAAATGTCTCCAGCCCTGAAC 2640  
 QY 2641 TTTGTTATTTTATGAGGATGAGGCTATACCTCAATTTTGTGACATGACCCAGAGA 2700  
 DB 2641 TTTGTTATTTTATGAGGATGAGGCTATACCTCAATTTTGTGACATGACCCAGAGA 2700  
 QY 2701 TATGTACAGTCAAGATTTACTCATAGAGCAATGAGACCTTTGACAGACATCTTGG 2760  
 DB 2701 TATGTACAGTCAAGATTTACTCATAGAGCAATGAGACCTTTGACAGACATCTTGG 2760  
 QY 2761 AGGGATTCGCAAGTGGAGCGCCCTGTTGATATCTATCTATCATCATTTTAAACAGCTGAT 2820  
 DB 2761 AGGGATTCGCAAGTGGAGCGCCCTGTTGATATCTATCTATCATCATTTTAAACAGCTGAT 2820  
 QY 2821 GGCCTTACCTTGGTGAAGTAAACCAATATATACATGATGCTGCTCAAGAGTGGCTTGA 2880  
 DB 2821 GGCCTTACCTTGGTGAAGTAAACCAATATATGACATGATGCTGCTCAAGAGTGGCTTGA 2880  
 QY 2881 ATATATCCCATGGGTTATCTGATGAGTACTGAGGTTTGAAGAGACTAGCCATATC 2940  
 DB 2881 ATATATCCCATGGGTTATCTGATGAGTACTGAGGTTTGAAGAGACTAGCCATATC 2940  
 QY 2941 TAGCATCTTACTGCTTATCTGCTTATGCTTGGGTTGGGTTAGGTAGATACCAA 3000  
 DB 2941 TAGCATCTTACTGCTTATCTGCTTATGCTTGGGTTGGGTTAGGTAGATACCAA 3000  
 QY 3001 TGAACACTTTCAGAGACCTTCTCTCTGCAAGTGTCTTAACTCTTACTAGAG 3060  
 DB 3001 TGAACACTTTCAGAGACCTTCTCTCTGCAAGTGTCTTAACTCTTACTAGAG 3060  
 QY 3061 GAGATTAATATTTTGCATTAATGAAGAAATTTCTAGTATATTAAGCAGGCTTTTAT 3120  
 DB 3061 GAGATTAATATTTTGCATTAATGAAGAAATTTCTAGTATATTAAGCAGGCTTTTAT 3120  
 QY 3121 TTTCTAAATATGATGATATGATTAATGTTAGATTAACGAATTTTGAATTTCCAG 3180  
 DB 3121 TTTCTAAATATGATGATATGATTAATGTTAGATTAACGAATTTTGAATTTCCAG 3180  
 QY 3181 AGAATATTTTAAAGTCTTATGATGAAATTAATCATCTTGTCTGATTTAAAAAAA 3240  
 DB 3181 AGAATATTTTAAAGTCTTATGATGAAATTAATCATCTTGTCTGATTTAAAAAAA 3240  
 QY 3241 AAA 3243  
 DB 3241 AAA 3243  
 RESULT 2  
 AF307339 3138 bp mRNA linear PRI 13-FEB-2001  
 LOCUS AF307339  
 DEFINITION Homo sapiens B aggressive lymphoma short isoform (BAL) mRNA,  
 complete cds.  
 ACCESSION AF307339  
 VERSION AF307339.1 GI:12751140  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 3138)  
 Aguiar R.C., Yakushiji J., Kharbanda S., Salgia R., Fletcher J.A.  
 and Shipp M.A. BAL is a novel risk-related gene in diffuse large B-cell lymphomas  
 that enhances cellular migration  
 Blood 96 (13), 4328-4334 (2000)  
 TITLE  
 JOURNAL MEDLINE 20563954  
 PUBMED 1110709  
 REFERENCE 2 (bases 1 to 3138)  
 Aguiar R.C.T., Yakushiji J., and Shipp M.A.  
 Direct Submission  
 TITLE Submitted (21-SEP-2000) Adult Oncology, Dana Farber Cancer  
 JOURNAL Institute, 44 Binney Street, Boston, MA 02115, USA

FEATURES  
source  
Location/Qualifiers  
1. .3138  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3q13-q21"  
1. .3138  
/gene="BAL"  
229. .2688  
/gene="BAL"  
/note="overexpressed in high risk diffuse large B-cell lymphoma"  
/codon\_start=1  
/product="B aggressive lymphoma short isoform"  
/protein\_id="AAK07559.1"  
/db\_xref="GI:12751141"  
/translation="MDPSWVAGAAAYNEKSENGALGENYSWQPIRHNPKILKNNR  
QICEVLQNKFGCTISTLVSYQEGNSKSLQVFRKMLPTRIELSWKDLTHAVDAVN  
AANEDLLHGGLALALVKAQGEIIEESKQFVARVYKVASGELIAYGAGRLPKQIIL  
AVGPRMENDKQCTGKLRALIVSLINYLKNTIKYVAPALSSGIFQFPLNCTK  
TIVETIRVLSLOGKPMNSLKEIHLVSNEDPTVAFKASEEILSGSELQOETPSNA  
MYNNLTLQVGHIEQMADVIVNSVNDHDLTVGVVASKILOQAGVEKMSIELATKA  
KOPORSQVLYVRKFNLFCKYIYHVMHSEPRPOLIKAMRECELEQONITSISF  
PALGTENMEIKETAEILFDEVLYFAKHVHQTLVKFVLEPTDLEIKAESSEMAK  
RSKMLSLNNYVSPQSTREKRENGLEAREALMLQFNEEMCEAHAWIORILSLNH  
HIENNNHILYLRKEHDIISQLOKTSVSITELISPRELEIEGARADLIEVNNIE  
DMICVQOEEMARKKERGLWSLQOMTIOOKTODEKEMIFLCKPVPLOEDQK  
OFKCGLOYLKYERIDNEVLMFAFORKKMEKILHROVSHRLEQOAVYQFNCNVCR  
VGFORNYSTPCDPKYGAGIYFTKNEKLNIAEKAKKISAALKLIYFEAEVLEPFCQGH  
PLNIVPPUSPGALDGDHSDVDNVSSPEFVIFSGMOALPQYILMTCTOEYVSDYSS  
GPMRPFQHPHWRFAFSQPD"

BASE COUNT 962 a 622 c 733 g 821 t  
ORIGIN

Query Match 93.2%; Score 3021.4; DB 9; Length 3138;  
Best Local Similarity 96.7%; Pred. No. 0;  
Matches 3137; Conservative 0; Mismatches 1; Indels 105; Gaps 1;

QY 1 GGGCTTCGTCCTGGGTGCTGACGTCGACGTCGCCGCCGAGGACTTGAAGTCT 60  
DB 1 GGGCTTCGTCCTGGGTGCTGACGTCGACGTCGCCGCCGAGGACTTGAAGTCT 60  
QY 61 GGAAGTAGCTCTCAGCTTCCTCTGACTCGGGGGCCGGAATTGTACACCCGACGAGA 120  
DB 61 GGAAGTAGCTCTCAGCTTCCTCTGACTCGGGGGCCGGAATTGTACACCCGACGAGA 120  
QY 121 GCGGGGACGCGCGCGGCGAGAGTGGCCACCATATCTGGAACCTACAGTCTATGCTTGA 180  
DB 121 GCGGGGACGCGCGCGGCGAGAGTGGCCACCATATCTGGAACCTACAGTCTATGCTTGA 180  
QY 181 AGCGCAAAAGGGAATTAACATTTAAAGACATCCCGCGGAGCTGGAGATGGAATTTCC 240  
DB 181 AGCGCAAAAGGGAATTAACATTTAAAGACATCCCGCGGAGCTGGAGATGGAATTTCC 240  
QY 241 ATGGTGGCCGAGCAGCAGCTTCAATGAATAATCAGATGAGATTACCTGCTCTCACTC 300  
DB 241 ATGGTGGCCGAGCAGCAGCTTCAATGAATAATCAGATGAGATTACCTGCTCTCACTC 300  
QY 301 TTGTTTCAGAAAGTCTTTGCTCAGATCTTTCCTCAGTGGAGAAAGGGAATACAGAAAG 360  
DB 274 ----- 273  
QY 361 TGTCTCCCTACAGAGTCTCAGAGACTGCTGCTTGGAGAAAACATTAAGTGGCAATTT 420  
DB 274 -----TAGAGACTGGTCTCTTGGAGAAAACATTAAGTGGCAATTT 315  
QY 421 CCCATTAAACCAATGACTTCAAAATTTTAAAAAATATAGCGTCAAGCTGTGAAGTC 480  
DB 316 CCCATTAAACCAATGACTTCAAAATTTTAAAAAATATAGCGTCAAGCTGTGAAGTC 375  
QY 481 CTCACAGATTAAGTTGGCTATCTCTACCTGCTCTCTCAGTTGAGGAAGCAACAGC 540  
DB 376 CTCACAGATTAAGTTGGCTATCTCTACCTGCTCTCTCAGTTGAGGAAGCAACAGC 435

QY 541 AAATCTCTGCAAGTGTTCAGAAAAATGCTGACTCTAGATAGAGTTATCAGTCTGGAAA 600  
DB 436 AAATCTCTGCAAGTGTTCAGAAAAATGCTGACTCTAGATAGAGTTATCAGTCTGGAAA 495  
QY 601 GATGACCTCAACACACATGCTGTTGATGCTGTGTGAATCAGCAATGAAGATCTTCTG 660  
DB 496 GATGACCTCAACACACATGCTGTTGATGCTGTGTGAATCAGCAATGAAGATCTTCTG 555  
QY 661 CATGGGGAGGCGCTGGCGCCCTGGCTGTAAAGCTGTGGATTTGAAATCCAAAGAGAG 720  
DB 556 CATGGGGAGGCGCTGGCGCCCTGGCTGTAAAGCTGTGGATTTGAAATCCAAAGAGAG 615  
QY 721 AGCAAAAGTTTGTGGCAGATATGTAAGTGTACCTGTGTAGATAGCTGTACAGGGA 780  
DB 616 AGCAAAAGTTTGTGGCAGATATGTAAGTGTACCTGTGTAGATAGCTGTACAGGGA 675  
QY 781 GCAAGGAGGCTTCCCTGCAAAACAGATCATCATGCTGTTGGCCCTCGGTGATGGAATG 840  
DB 676 GCAAGGAGGCTTCCCTGCAAAACAGATCATCATGCTGTTGGCCCTCGGTGATGGAATG 735  
QY 841 GATTAACAGGAGTACTGGAAGCTGACAGAGGCCATTGTAAGTATTCGAATATGTC 900  
DB 736 GATTAACAGGAGTACTGGAAGCTGACAGAGGCCATTGTAAGTATTCGAATATGTC 795  
QY 901 ATCTATAAATATCTCATTAAGACAGTAGCAATTCCAGCTTGAAGCTTGGGATTTT 960  
DB 796 ATCTATAAATATCTCATTAAGACAGTAGCAATTCCAGCTTGAAGCTTGGGATTTT 855  
QY 961 CAGTTCCTCTGAAATTTGTGTACAAAGACTATGTAAGACTATCCGGGTTAGTTGCAA 1020  
DB 856 CAGTTCCTCTGAAATTTGTGTACAAAGACTATGTAAGACTATCCGGGTTAGTTGCAA 915  
QY 1021 GGAAGCCCATGATGATATTTGAAGAAATTCACCTGTGACCATGAGACCCCTACT 1080  
DB 916 GGAAGCCCATGATGATATTTGAAGAAATTCACCTGTGACCATGAGACCCCTACT 975  
QY 1081 GTTGCTGCTCTTTAAAGCTGCTTCAAGATTCATCTAGGAAAGTGAAGTGGGACAAAGAA 1140  
DB 976 GTTGCTGCTCTTTAAAGCTGCTTCAAGATTCATCTAGGAAAGTGAAGTGGGACAAAGAA 1035  
QY 1141 ACCACCCCTCTTCAATGCAATGCTGTAAGACACCTGACCTCCAGATTTGCCAGGCG 1200  
DB 1036 ACCACCCCTCTTCAATGCAATGCTGTAAGACACCTGACCTCCAGATTTGCCAGGCG 1095  
QY 1201 CACATTGAATGGAGACAGCAGATGTAATGTTAATCTCTAAACCACATGATATTTACA 1260  
DB 1096 CACATTGAATGGAGACAGCAGATGTAATGTTAATCTCTAAACCACATGATATTTACA 1155  
QY 1261 GTTGACCTGTGGCAAAAGTCAATTCACAAACAGCAGAGATTGAATGAATGGAATTT 1320  
DB 1156 GTTGACCTGTGGCAAAAGTCAATTCACAAACAGCAGAGATTGAATGAATGGAATTT 1215  
QY 1321 CTTGCCAAAGAGCTTAACAGTTTCAACGGTCCAGTTGCTACTGTCACAAAAGATTT 1380  
DB 1216 CTTGCCAAAGAGCTTAACAGTTTCAACGGTCCAGTTGCTACTGTCACAAAAGATTT 1275  
QY 1381 AACTGTTCTGTAAATATATATACATGCTGTGGATTCCAGAAATTTCTTAACCTCAG 1440  
DB 1276 AACTGTTCTGTAAATATATATACATGCTGTGGATTCCAGAAATTTCTTAACCTCAG 1335  
QY 1441 ATATTAACATCAATGAAGAGTGTGGAATAATGATGAGCAAAATATTAACCTTC 1500  
DB 1336 ATATTAACATCAATGAAGAGTGTGGAATAATGATGAGCAAAATATTAACCTTC 1395  
QY 1501 ATTTCTTCTCTGCTGCTGGAGCTGGAACATGGAATAAAGAGAAACAGCAGAGAG 1560  
DB 1396 ATTTCTTCTCTGCTGCTGGAGCTGGAACATGGAATAAAGAGAAACAGCAGAGAG 1455  
QY 1561 ATTTTGTGATGAAGTTTAACTTTGCCAAGACCATGTATAAACACCGATTAACTGTA 1620  
DB 1456 ATTTTGTGATGAAGTTTAACTTTGCCAAGACCATGTATAAACACCGATTAACTGTA 1515

DB	Accession	Source	Sequence	Start	End	Strand	Product
Db	2556	TATGTACAGTCAACAAGATTACTACATCAGACCAAGAAAGAACCCCTTTGGACACATCTCTGG					2655
QY	2761	AGGGATTTCGCAAGAGCGAGCCCTGGTGAATTAATCTCAATCATATTTTAAACAGCTGTAT					2820
Db	2656	AGGGATTTCGCAAGAGCGAGCCCTGGTGAATTAATCTCAATCATATTTTAAACAGCTGTAT					2715
QY	2821	GGCCTTACCTTGGGTGGAACATAACCAATATATGACCATCGATGCGTCAAAAGTGGCTTGA					2880
Db	2716	GGCCTTACCTTGGGTGGAACATAACCAATATATGACCATCGATGCGTCAAAAGTGGCTTGA					2775
QY	2881	ATATATCCCATGGGTATCTGTATGAGACTGACTGGGTATTTGAAAGAGCTATGCCACATAC					2940
Db	2776	ATATATCCCATGGGTATCTGTATGAGACTGACTGGGTATTTGAAAGAGCTATGCCACATAC					2835
QY	2941	TAGCATCTTAGTGCCTTATATCTGTATATGCTTGGGGTGGGTAGTATACCAAA					3000
Db	2836	TAGCATCTTAGTGCCTTATATCTGTATATGCTTGGGGTGGGTAGTATACCAAA					2895
QY	3001	TGAACACTTTCAGAGACCTCCCTCCCTGTCAGTGTGTCCTTAATCTCCTTACTAGAG					3060
Db	2896	TGAACACTTTCAGAGACCTCCCTCCCTGTCAGTGTGTCCTTAATCTCCTTACTAGAG					2955
QY	3061	GAGATTAATATTTTGCATATATAATGAGAAATTTTCTGTATATATACGACGCTTTAT					3120
Db	2956	GAGATTAATATTTTGCATATATAATGAGAAATTTTCTGTATATATACGACGCTTTAT					3015
QY	3121	TTTCTTAAATATGAT					3180
Db	3016	TTTCTTAAATATGAT					3075
QY	3181	AGATATATATATAAGTCTTATAGTATGAAATTAATCATCTTTGTCTGATTAATAAAAAA					3240
Db	3076	AGATATATATATAAGTCTTATAGTATGAAATTAATCATCTTTGTCTGATTAATAAAAAA					3135
QY	3241	AAA 3243					
Db	3136	AAA 3138					

PALGTGMEIKKETAEILFDEVILFPAHDVHKHOLTVKFEVLPDLEIYKAFSSBMAK  
RSKMLSLNNYSVPPOSTBREKRENEGLEASPAIINLISPEELEYEYEAHAIJORI.LS.IONH  
HI.IENNHLIYLRKEDHILISOLOKTSVSIPIIISPEELEYEYEAHAIJORI.LS.IONH  
DMICKVOEEMARKREKGLMRSLIOWTIOOKOTDOKMKNIIIFLKPVPPOEILDOOK  
OPEKGLQVLYEKERIDNEVLMAPORKKRMEBEKJHRPVSRLRYEQOYPYOCNVYCR  
V6FQMYSTPCDPEKAGIYFTFKLNKLNAEBKAKIISADKLIYVEAEVLTGFCOGH  
PLNIVPLSLPGALDHDSDVVDNVSPTFVILFSQMOAIPOLIMTCTOEYVOSODYSS  
GPMRPFACHPHMFGSGSPVD\*  
BASE COUNT 932 a 589 c 696 g 799 t  
ORIGIN

Query Match 87.7%; Score 2842.6; DB 6; Length 3016;  
Best Local Similarity 96.3%; Pred. No. 0;  
Matches 2963; Conservative 0; Mismatches 9; Indels 105; Gaps 1;

QY 151 CAPATCTGGAACCTACAGTCTATGCTTGTGAGCGCAAAAGGGAATTAACATTAAAGACT 210  
DB 45 CAGGCTGGAACTACAGTCTATGCTTGAAGCGCAAAAGGGAATTAACATTAAAGACT 104  
QY 211 CCCCCGGGACCTGAGAGATGACCTTTTCCATGTTGGCCGGAGCAGCAGCTTACAATGAA 270  
DB 105 CCCCCGGGAGACCTGAGAGATGACCTTTTCCATGTTGGCCGGAGCAGCAGCTTACAATGAA 164  
QY 271 AATACAGTAGATTACCTCGCTCTCAGTCTTGTTCAGAAAGTCTTGTCTCAGATCTTT 330  
DB 165 AAA----- 167  
QY 331 CCTCAGTGAGAAAGGGAATACAGAAAGATGTCCTCCCTACAAAGTGTCCAGAGACTGGT 390  
DB 168 ----- -TCAGAGACTGGT 179  
QY 391 GCTCTTGGAGAAAACATATAGTGGCAAAATTCACATTAACACATGACTTCAAAATTTTA 450  
DB 180 GCTCTTGGAGAAAACATATAGTGGCAAAATTCACATTAACACATGACTTCAAAATTTTA 239  
QY 451 AAAAATAATAGACGCTACGCTGTGTGAAGTCCCTCAGAAATTAAGTTGGCTGTATCTACC 510  
DB 240 AAAAATAATAGACGCTACGCTGTGTGAAGTCCCTCAGAAATTAAGTTGGCTGTATCTACC 299  
QY 511 CTGTGCTCTCCAGTTCAGGAAGGCAACAGCAAAATCTGTCAAGTGTTCAGAAAAATGCTG 570  
DB 300 CTGTGCTCTCCAGTTCAGGAAGGCAACAGCAAAATCTGTCAAGTGTTCAGAAAAATGCTG 359  
QY 571 ACTCTAGAGATAGATTATCAGTCTGTGAAAGATGACCTCACCACATGCTGTGTGATGCT 630  
DB 360 ACTCTAGAGATAGATTATCAGTCTGTGAAAGATGACCTCACCACATGCTGTGTGATGCT 419  
QY 631 GTGTGTAATGCAAGCCATGAAAGATCTTCTGCATGGGGAGGCTGGCCCTGCGCTGCTGTA 690  
DB 420 GTGTGTAATGCAAGCCATGAAAGATCTTCTGCATGGGGAGGCTGGCCCTGCGCTGCTGTA 479  
QY 691 AAAGGTGTGATTTGAATTCACAAGAGAGACAACAGTGTGTTGCCAGATATGTTAA 750  
DB 480 AAAGGTGTGATTTGAATTCACAAGAGAGACAACAGTGTGTTGCCAGATATGTTAA 539  
QY 751 GTGTCACTGCTGTAGATAGCTGTCAAGGAGCAGGAGGCTTCCCTGCAACAAGATCATC 810  
DB 540 GTGTCACTGCTGTAGATAGCTGTCAAGGAGCAGGAGGCTTCCCTGCAACAAGATCATC 599  
QY 811 CATGCTGTGGGCTCTGGGTGATGGAATGGGATAAACAGGAGTGTACTGTGAAAGCTGCGAG 870  
DB 600 CATGCTGTGGGCTCTGGGTGATGGAATGGGATAAACAGGAGTGTACTGTGAAAGCTGCGAG 659  
QY 871 AAGGCCATTGTAGTATTTGCAATTTGTCATCTATATAAATACACATTAAGACAGTA 930  
DB 660 AAGGCCATTGTAGTATTTGCAATTTGTCATCTATATAAATACACATTAAGACAGTA 719  
QY 931 GCAATTTCCAGCCTTGAGCTCTGGGATTTTTCAGTTCCTCTGTAATTTGTGTACAAGACT 990  
DB 720 GCAATTTCCAGCCTTGAGCTCTGGGATTTTTCAGTTCCTCTGTAATTTGTGTACAAGACT 779  
QY 991 ATTGTAGACATATCCGGGTTAGTTTGAAGGAACCAATGATGATTAATTTGAAAGAA 1050

DB 780 ATTGTAGACATATCCGGGTTAGTTTGCAGGGAACCAATGATGATTAATTTGAAAGAA 839  
QY 1051 ATTCACTGCTGTGAGCAATGAGACCCCTACTGTTGCTGCCCTTTAAAGCTGCTCAGAAATTC 1110  
DB 840 ATTCACTGCTGTGAGCAATGAGACCCCTACTGTTGCTGCCCTTTAAAGCTGCTCAGAAATTC 899  
QY 1111 ATTCCTAGGAAGAGTAGCTGGGACAAAGAACCCCTCTTCTTCAATGCAATGCTGCTG 1170  
DB 900 ATTCCTAGGAAGAGTAGCTGGGACAAAGAACCCCTCTTCTTCAATGCAATGCTGCTG 959  
QY 1171 AACACCTGACCTCCAGATTTGTCAGGGCCACATGTAATGGCAGACGGCAGATGTAAT 1230  
DB 960 AACACCTGACCTCCAGATTTGTCAGGGCCACATGTAATGGCAGACGGCAGATGTAAT 1019  
QY 1231 GTTAAATTTGTAAACCCACATGATTAATACAGTTGAGCTGTGGCAAGTCAATTTACAA 1290  
DB 1020 GTTAAATTTGTAAACCCACATGATTAATACAGTTGAGCTGTGGCAAGTCAATTTACAA 1079  
QY 1291 CAAGCAGAGGTTGAAATGGAATTCGAATTTCTGCCAAGAGCTTAACAGTTTCAACGG 1350  
DB 1080 CAAGCAGAGGTTGAAATGGAATTCGAATTTCTGCCAAGAGCTTAACAGTTTCAACGG 1139  
QY 1351 TCCCATGTTGCTAGCTGTCAACAAAGATTTAACTGTCTGTAAATATATATACATGTA 1410  
DB 1140 TCCCATGTTGCTAGCTGTCAACAAAGATTTAACTGTCTGTAAATATATATACATGTA 1199  
QY 1411 CTGTGCAATTCAGAAATTTCTTAACCTCAGATTTAAACATGCAATGAAAGAGTGTG 1470  
DB 1200 CTGTGCAATTCAGAAATTTCTTAACCTCAGATTTAAACATGCAATGAAAGAGTGTG 1259  
QY 1471 GAAAAATGATGAGCAAAATATACCTTCATTTCTGTGCTGCTGGAGCTGGAAC 1530  
DB 1260 GAAAAATGATGAGCAAAATATACCTTCATTTCTGTGCTGCTGGAGCTGGAAC 1319  
QY 1531 ATGGAATTAAGAGGAAGACAGAGAGATTTGTTGATGAATTTAACTTAACATTTGGC 1590  
DB 1320 ATGGAATTAAGAGGAAGACAGAGAGATTTGTTGATGAATTTAACTTAACATTTGGC 1379  
QY 1591 AAAGACCATGTAAACACACAGTAACTGTAAATTTGTATGATCTTTCACACAGATTTGGAG 1650  
DB 1380 AAAGACCATGTAAACACACAGTAACTGTAAATTTGTATGATCTTTCACACAGATTTGGAG 1439  
QY 1651 ATATATTAAGGCTTTTCAGTTCTGAATGCAAAAGAGTCCAAAGATGCTGAATTTGAACAAT 1710  
DB 1440 ATATATTAAGGCTTTTCAGTTCTGAATGCAAAAGAGTCCAAAGATGCTGAATTTGAACAAT 1499  
QY 1711 TACAGTGTCCCCAGTCAACACAGAGAGAGAAAGAAATGGGCTTAAGCTAGATGCT 1770  
DB 1500 TACAGTGTCCCCAGTCAACACAGAGAGAGAAAGAAATGGGCTTAAGCTAGATGCT 1559  
QY 1771 CCTGCCATCAATCTGATGGGATTTCAACGTGAAGAGATGATGAGGCCACACGATGATC 1830  
DB 1560 CCTGCCATCAATCTGATGGGATTTCAACGTGAAGAGATGATGAGGCCACACGATGATC 1619  
QY 1831 CAAGAAATCTGAGTCTCAGAAACCAACATCATTAATGAGAATTAATCATATTTGTAACCTT 1890  
DB 1620 CAAGAAATCTGAGTCTCAGAAACCAACATCATTAATGAGAATTAATCATATTTGTAACCTT 1679  
QY 1891 GGGAGAAAGAACATGACATTTTGTCTCAGCTTCAGAAAACCTTCAAGTGTCTCCATCACA 1950  
DB 1680 GGGAGAAAGAACATGACATTTTGTCTCAGCTTCAGAAAACCTTCAAGTGTCTCCATCACA 1739  
QY 1951 GAAATTTATCAGCCACAGAGAGACAGATTTAGAGATTTGAAGAGCCCGGCTCAGCTCAT 2010  
DB 1740 GAAATTTATCAGCCACAGAGAGACAGATTTAGAGATTTGAAGAGCCCGGCTCAGCTCAT 1799  
QY 2011 GAGGTGTTATGAAACATTTGAAGATATGTTGTAAAGTACAGAGGAAATGGCAAGAAA 2070  
DB 1800 GAGGTGTTATGAAACATTTGAAGATATGTTGTAAAGTACAGAGGAAATGGCAAGAAA 1859  
QY 2071 AAGGAGCAGGCGCTTTGGGCTGCTGTAGAGACAGTGAATTCAGACACAAAAAACCCAA 2130  
DB 1860 AAGGAGCAGGCGCTTTGGGCTGCTGTAGAGACAGTGAATTCAGACACAAAAAACCCAA 1919

QY 2131 GAGGAATGAAGAAATATCATATTTCTGAAATGCTGCTGCCACTCAGAGACTT 2190  
|||||  
Db 1920 GAGGAATGAAGAAATATCATATTTCTGAAATGCTGCTGCCACTCAGAGACTT 1919  
QY 2191 CTAGATCAAAAAGAAACAGTTGAAAAATGTTGCGAGTTCTTAAAGTGCGAGAGATA 2250  
|||||  
Db 1980 CTAGATCAAAAAGAAACAGTTGAAAAATGTTGCGAGTTCTTAAAGTGCGAGAGATA 2039  
QY 2251 GACAATGAGTCTCTATAGGCTGCTTCAAGAAAGAAAGAAATGATGGAAGAAAACTG 2310  
|||||  
Db 2040 GACAATGAGTCTCTATAGGCTGCTTCAAGAAAGAAAGAAATGATGGAAGAAAACTG 2099  
QY 2311 CACAGGCAACCTGTGAGCCATGAGGCTGTTGAGAGATGCCATCCAGTCTGCAATGTG 2370  
|||||  
Db 2100 CACAGGCAACCTGTGAGCCATGAGGCTGTTGAGAGATGCCATCCAGTCTGCAATGTG 2159  
QY 2371 GTATGCAAGTGTGGCTTCAAGAGATGACTCGACACTTGGCATGCCAAATAGGAGCT 2430  
|||||  
Db 2160 GTATGCAAGTGTGGCTTCAAGAGATGACTCGACACTTGGCATGCCAAATAGGAGCT 2219  
QY 2431 GGCATATCTTCAACCAAGAACTTCAAAAACCTGGCAGAGAAAGGCCAAGAAATCTGCT 2490  
|||||  
Db 2220 GGCATATCTTCAACCAAGAACTTCAAAAACCTGGCAGAGAAAGGCCAAGAAATCTGCT 2279  
QY 2491 GCAGATAGGCTGATCTATGTTGTTGAGGCTGAGTACTCAACAGGCTCTTCTGCCAGGA 2550  
|||||  
Db 2280 GCAGATAGGCTGATCTATGTTGTTGAGGCTGAGTACTCAACAGGCTCTTCTGCCAGGA 2339  
QY 2551 CATCCGTTAAATATTTGTTCCCAACCACTAGTCCAGTCTGAGCTATAGATGTCATGACAGT 2610  
|||||  
Db 2340 CATCCGTTAAATATTTGTTCCCAACCACTAGTCCAGTCTGAGCTATAGATGTCATGACAGT 2399  
QY 2611 GTGTTGACATATGCTCCAGCCCTGAAACCTTGTATTTTATGTCATGACAGCTATA 2670  
|||||  
Db 2400 GTGTTGACATATGCTCCAGCCCTGAAACCTTGTATTTTATGTCATGACAGCTATA 2459  
QY 2671 CCTCAGATTTTGGAGCATGACCCAGAGATATGTACAGTACAGAGATTTCTCATCGA 2730  
|||||  
Db 2460 CCTCAGATTTTGGAGCATGACCCAGAGATATGTACAGTACAGAGATTTCTCATCGA 2519  
QY 2731 CCAATGAGACCTTTGGCAGACATCTTTGAGAGGGATTTGCAAGTGGCAGCCCTGTGAT 2790  
|||||  
Db 2520 CCAATGAGACCTTTGGCAGACATCTTTGAGAGGGATTTGCAAGTGGCAGCCCTGTGAT 2579  
QY 2791 TAATCTCAATCATTTTAAACAGCTGTATGCTTACCTTGGTGAACCTAACCAATTA 2850  
|||||  
Db 2580 TAATCTCAATCATTTTAAACAGCTGTATGCTTACCTTGGTGAACCTAACCAATTA 2639  
QY 2851 TGACCATGATGGCTCAAGAGTGGCTTGAATATATCCATGGGTATTCGTATGAGCTG 2910  
|||||  
Db 2640 TGACCATGATGGCTCAAGAGTGGCTTGAATATATCCATGGGTATTCGTATGAGCTG 2699  
QY 2911 ACTGGGTTATTTGAAGAGACTAGCCACATACTAGCATCTTATGCTTATCTCTTAT 2970  
|||||  
Db 2700 ACTGGGTTATTTGAAGAGACTAGCCACATPACTAGCATCTTATGCTTATCTCTTAT 2759  
QY 2971 GTCTTGGGGTGGGGTGGAGTACCAATGAACACTTTCAGAGCTTCTCTCTCT 3030  
|||||  
Db 2760 GTCTTGGGGTGGGGTGGAGTACCAATGAACACTTTCAGAGCTTCTCTCTCTCT 2819  
QY 3031 GCACTTCTTCTTAAATCTCTTACTAGAGAGATTAATTTTGGCATPATAATGAAGAA 3090  
|||||  
Db 2820 GCACTTCTTCTTAAATCTCTTACTAGAGAGATTAATTTTGGCATPATAATGAAGAA 2879  
QY 3091 TTTTCTAGATATATACGAGGCTTTTATTTTCTAAATGATGATATATAAATGCTT 3150  
|||||  
Db 2880 TTTTCTAGATATATACGAGGCTTTTATTTCTAAATGATGATATATAAATGCTT 2939  
QY 3151 AGGATTAACAGATATTTTATGATTTTCCAGAGATATATATAAGTCTTATAGTATGAA 3210  
|||||  
Db 2940 AGGATTAACAGATATTTTATGATTTTCCAGAGATATATATAAGTCTTATAGTATGAA 2999

QY 3211 ATAAATCATCTTTGCT 3227  
|||||  
Db 3000 ATAAATCATCTTTGCT 3016

RESULT 4  
AX274946 2598 bp DNA linear PAT 29-OCT-2001  
LOCUS  
DEFINITION Sequence 211 from Patent WO0172777.  
ACCESSION AX274946  
VERSION AX274946.1 GI:16547578  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 Hillman, J.L., Baughn, M.R., Yue, H., Lal, P., Lu, D.A., Patterson, C., Azimzai, Y., Bandman, O., Tang, Y.T., Mathur, P., Shah, P., Au-Young, J. and Reddy, R.  
Transcription factors  
JOURNAL Patent: WO 017277-A 211 04-OCT-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
source location/Qualifiers  
1..2598  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 557394CB1"

BASE COUNT 823 a 491 c 587 g 697 t

ORIGIN

Query Match 78.4%; Score 2543.8; DB 6; Length 2598;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 2545; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 696 TGGTGCATTTTGAATTCAGAGAGAGACAAACAGTTTGTTCACAGATATGTAAGTGTG 755  
|||||  
Db 52 TGGTGCATTTTGAATTCAGAGAGAGACAAACAGTTTGTTCACAGATATGTAAGTGTG 111  
QY 756 AGCTGTGAGATAGCTGTACAGGAGAGAGAGGCTTCCCTGCMAACAGATCATCATGC 815  
|||||  
Db 112 AGCTGTGAGATAGCTGTACAGGAGAGAGAGGCTTCCCTGCMAACAGATCATCATGC 171  
QY 816 TGGTGGGCTGTGGAGTGAATGGAAATGAGAGAGTGTCTGGAAGCTGCAGAGGGC 875  
|||||  
Db 172 TGGTGGGCTGTGGAGTGAATGGAAATGAGAGAGTGTCTGGAAGCTGCAGAGGGC 231  
QY 876 CATTTAAGTATTTGTAATTTATGTCATCTATAAAATCTCACATTAAGACATAGCAAT 935  
|||||  
Db 232 CATTTAAGTATTTGTAATTTATGTCATCTATAAAATCTCACATTAAGACATAGCAAT 291  
QY 936 TCCAGCCTTGAGCTGTGGGATTTTTCAGTTCCTCTGTGATTTGTGTCAAGAGACTATTGT 995  
|||||  
Db 292 TCCAGCCTTGAGCTGTGGGATTTTTCAGTTCCTCTGTGATTTGTGTCAAGAGACTATTGT 351  
QY 996 AGAGACTATCCGGTATGTTTTCAGAGGAAGCCATGATGTGTAATTTGAAGAAATTCA 1055  
|||||  
Db 352 AGAGACTATCCGGTATGTTTTCAGAGGAAGCCATGATGTGTAATTTGAAGAAATTCA 411  
QY 1056 CCTGGTGAGCAATAGAGACCTTACTGTTGCTGCCCTTTAAACCTCTCAGAAATTCATCCT 1115  
|||||  
Db 412 CCTGGTGAGCAATAGAGACCTTACTGTTGCTGCCCTTTAAACCTCTCAGAAATTCATCCT 471  
QY 1116 AGGAGAGAGTAGCTGGGACAGAAACACCCCTTCTTTCATATGATGTCGTGAACAA 1175  
|||||  
Db 472 AGGAGAGAGTAGCTGGGACAGAAACACCCCTTCTTTCATATGATGTCGTGAACAA 531  
QY 1176 CCTGACCTCCAGATTTGTCAGGGCCACATTTGAATGGACGAGCGAGATGTAATTTGTTAA 1235  
|||||  
Db 532 CCTGACCTCCAGATTTGTCAGGGCCACATTTGAATGGACGAGCGAGATGTAATTTGTTAA 591  
QY 1236 TTCTGTAAACCCACATGATATTTACAGTTTGAACCTGTGGCAAGTCAATTTCAACAAGC 1295  
|||||

Db 592 TTCTTAACCCACATGATATTACAGTTGGACCTGGCCAAAGTCATTTCTCAACAACG 651  
Qy 1296 AGAGTTGAAATGAATCGGAATTTCTTGCCACAAAGGCTAAACAGTTTCAACGGTCCA 1355  
Db 652 AGAGTTGAAATGAATCGGAATTTCTTGCCACAAAGGCTAAACAGTTTCAACGGTCCA 711  
Qy 1356 GTTGGTACTGGTCACAAAAGATTTAACTGTTCTGTAATATATATACATGTACTGTCG 1415  
Db 712 GTTGGTACTGGTCACAAAAGATTTAACTGTTCTGTAATATATATACATGTACTGTCG 771  
Qy 1416 GCATTCCAGAAATTCCTAAACCTCAGATATTAACATGCAAGAAAGGATGTTGGAAAA 1475  
Db 772 GCATTCCAGAAATTCCTAAACCTCAGATATTAACATGCAAGAAAGGATGTTGGAAAA 831  
Qy 1476 ATGCATTGAGCAAAATATATACCTTCATTTCTCTGCTTGGGACTGGAACATGGA 1535  
Db 832 ATGCATTGAGCAAAATATATACCTTCATTTCTCTGCTTGGGACTGGAACATGGA 891  
Qy 1536 AATAAAGAAGAAACAGCAGCAGAGATTTGTTGATGAAGTTTAACTTTGCCAAGA 1595  
Db 892 AATAAAGAAGAAACAGCAGCAGATTTGTTGATGAAGTTTAACTTTGCCAAGA 951  
Qy 1596 CCATGTAAACACCGATTACGTAAATTTTGATCTTCCACAGATTTGGAGATATA 1655  
Db 952 CCATGTAAACACCGATTACGTAAATTTTGATCTTCCACAGATTTGGAGATATA 1011  
Qy 1656 TAAGCTTTCACTTGTGAATGGAAGAGTCCAAAGATGCTAGTTGACAAATTCAG 1715  
Db 1012 TAAGCTTTCACTTGTGAATGGAAGAGTCCAAAGATGCTAGTTGACAAATTCAG 1071  
Qy 1716 TGTCCCCAGTCAACCCAGAGAGAGAAAAAGAAAAATGGGCTTGAAGCTAGATCTCTGC 1775  
Db 1072 TGTCCCCAGTCAACCCAGAGAGAGAAAAAGAAAAATGGGCTTGAAGCTAGATCTCTGC 1131  
Qy 1776 CATCATCTGATGGGATTAAAGTGAAGATGATGAGGCCCGACCGATCCAAAG 1835  
Db 1132 CATCATCTGATGGGATTAAAGTGAAGATGATGAGGCCCGACCGATCCAAAG 1191  
Qy 1836 AATCTGAGTCTCCAGAAACACACATCATTTGAGAAATATCATATTTCTTACCTTGGAG 1895  
Db 1192 AATCTGAGTCTCCAGAAACACACATCATTTGAGAAATATCATATTTCTTACCTTGGAG 1251  
Qy 1896 AAAGGACATGACATTTTGTCTCAGCTTCAGAAAACCTTCAAGTGTCTCATCAGCAAT 1955  
Db 1252 AAAGGACATGACATTTTGTCTCAGCTTCAGAAAACCTTCAAGTGTCTCATCAGCAAT 1311  
Qy 1956 TATCAGCCGAGAGAGACAGATTTAGATTTGAAGAGCCGGGCTGACCTCATTTAGAT 2015  
Db 1312 TATCAGCCGAGAGAGACAGATTTAGATTTGAAGAGCCGGGCTGACCTCATTTAGAT 1371  
Qy 2016 GGTATGAAACATTTGAAGATATGCTTTGTAAATACAGAGAAATGCGAAGAAAAAGA 2075  
Db 1372 GGTATGAAACATTTGAAGATATGCTTTGTAAATACAGAGAAATGCGAAGAAAAAGA 1431  
Qy 2076 GCGAGGCTTTGGCGCTGTTAGACAGTGAATTCAGCAACAAAAAACCAGAGCA 2135  
Db 1432 GCGAGGCTTTGGCGCTGTTAGACAGTGAATTCAGCAACAAAAAACCAGAGCA 1491  
Qy 2136 AATGAAAGAAATATCATATTTCTGAAATGTCCTGCTCCACATCAGACACTTCTA 2195  
Db 1492 AATGAAAGAAATATCATATTTCTGAAATGTCCTGCTCCACATCAGACACTTCTA 1551  
Qy 2196 TCAAAAGAAACAGTTGAAAAATGTGTTGAGGTTCTAAAGGTGAGAGATAGACAA 2255  
Db 1552 TCAAAAGAAACAGTTGAAAAATGTGTTGAGGTTCTAAAGGTGAGAGATAGACAA 1611  
Qy 2256 TGAGTCTCTTATGCTGCTTTCAAGAAAGAAATGATGGAAGAAAACTGCACAG 2315  
Db 1612 TGAGTCTCTTATGCTGCTTTCAAGAAAGAAATGATGGAAGAAAACTGCACAG 1671  
Qy 2316 GCAACCTGAGACCATAGGCTTTTCAGACAACTCCATCCAGTCTTGCATAGTGTATG 2375  
Db 1672 GCAACCTGAGACCATAGGCTTTTCAGACAACTCCATCCAGTCTTGCATAGTGTATG 1731

Qy 2376 CAGAGTTGGCTTTCAAGAAATGTACTGCACACCTTGGCATCCAAATACGAGCTGGCAT 2435  
Db 1732 CAGAGTTGGCTTTCAAGAAATGTACTGCACACCTTGGCATCCAAATACGAGCTGGCAT 1791  
Qy 2436 ATACCTTCACCAAGAACCTTCAAAACCCTGGCAGAGAAAGCCCAAGAAATCTGCTGCAGA 2495  
Db 1792 ATACCTTCACCAAGAACCTTCAAAACCCTGGCAGAGAAAGCCCAAGAAATCTGCTGCAGA 1851  
Qy 2496 TAAAGTATCTATGTGTTGAGGCTGAAGTACTCAGAGCTTCTCTGACAGGACATCC 2555  
Db 1852 TAAAGTATCTATGTGTTGAGGCTGAAGTACTCAGAGCTTCTCTGACAGGACATCC 1911  
Qy 2556 GTTAAATATTTGTTCCCCACACATGATCTTGAGCTATAGATGCTATGACATGTGGT 2615  
Db 1912 GTTAAATATTTGTTCCCCACACATGATCTTGAGCTATAGATGCTATGACATGTGGT 1971  
Qy 2616 TGAACAATGCTCCAGCCCTGAACCTTGTATTTTATGTCATGACAGCTATACCTCA 2675  
Db 1972 TGAACAATGCTCCAGCCCTGAACCTTGTATTTTATGTCATGACAGCTATACCTCA 2031  
Qy 2676 GTATTTGTGACATGACACCCAGAAATATGTACAGTCACAGATTAATCTCAGACCAAT 2735  
Db 2032 GTATTTGTGACATGACACCCAGAAATATGTACAGTCACAGATTAATCTCAGACCAAT 2091  
Qy 2736 GAGACCTTTGACACAGCATCTTTGAGGGGATTCGCAAGTGCAGCCCTGTTGATTAATC 2795  
Db 2092 GAGACCTTTGACACAGCATCTTTGAGGGGATTCGCAAGTGCAGCCCTGTTGATTAATC 2151  
Qy 2796 TCTACATCATTTTAAACAGCTGATGCGCTTACCTTGGGTGAACATTAACAAATTAATGAC 2855  
Db 2152 TCTACATCATTTTAAACAGCTGATGCGCTTACCTTGGGTGAACATTAACAAATTAATGAC 2211  
Qy 2856 ATCGATGCTCAAGAGTGGCTTGAATATATCCATGGGTATCTGTATGACTGACTGG 2915  
Db 2212 ATCGATGCTCAAGAGTGGCTTGAATATATCCATGGGTATCTGTATGACTGACTGG 2271  
Qy 2916 GTTATTTGAAGAGATAGCCACATACATCTAGATTTAGTGGCTTTATCTGCTTTATGCTTT 2975  
Db 2272 GTTATTTGAAGAGATAGCCACATACATCTAGATTTAGTGGCTTTATCTGCTTTATGCTTT 2331  
Qy 2976 GGGGTTGGGGTATGATATACCAATGAACACTTCAGAGACTTCCTTCCCTTGGCACT 3035  
Db 2332 GGGGTTGGGGTATGATATACCAATGAACACTTCAGAGACTTCCTTCCCTTGGCACT 2391  
Qy 3036 TGTCTTTATCTCTCTTACAGAGATTAATTTTGCATATATGAAGAAATTTT 3095  
Db 2392 TGTCTTTATCTCTCTTACAGAGATTAATTTTGCATATATGAAGAAATTTT 2451  
Qy 3096 CTAGTATATACGACAGGCTTTTATTTCTAAATGATGATAGTAAATGTTAGAT 3155  
Db 2452 CTAGTATATACGACAGGCTTTTATTTCTAAATGATGATAGTAAATGTTAGAT 2511  
Qy 3156 AACAGATATTTTATGATTTTCCAGAGATATTTATTAAGTCTTTTATGATGATGATGAT 3215  
Db 2512 AACAGATATTTTATGATTTTCCAGAGATATTTATTAAGTCTTTTATGATGATGATGAT 2571  
Qy 3216 TCAATCTTGTCTGATTTAAAAAATAA 3242  
Db 2572 TCAATCTTGTCTGATTTAAAAAATAA 2598

RESULT 5  
HSM803023 1442 bp mRNA linear PRI 20-MAR-2002  
LOCUS HSM803023  
DEFINITION Homo sapiens mRNA; cDNA DKFZp76111617 (from clone DKFZp76111617).  
ACCESSION AL713679  
VERSION AL713679.1 GI:19584378  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1442)  
AUTHORS Ansoore,W., Mirkner,U., Nemes,H.W., Well,B. and Wiemann,S.  
TITLE Direct Submission  
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152  
Martinsried, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
This clone (DKFZp761I1617) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
information about the clone and the sequencing project is available  
at <http://mips.gsf.de/proj/cDNA/>.  
FEATURES  
Location/Qualifiers  
1..1442  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="3"  
/clone="DKFZp761I1617"  
/cissue\_type="amygdala"  
/cissue\_lib="761 (synonym: hamy2). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
1..1442  
/gene="DKFZp761I1617"  
<1..993  
/gene="DKFZp761I1617"  
/note="BAL"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD28439.1"  
/db\_xref="GI:19584379"  
/translation="EMEYEAHWIOTRILISLNHIIENNHLIYLRKEHDILOLOKT  
SSVITTEIISGRTELEIGARADLIEVNNIIEMLCKVDEMARKKERGLMSIGOM  
TIOOKTODEKKNIEIFLKCPPTDLDQKOEKGLVLEKEDINVLMAFO  
RKKKMEKKIASHROVSHRLEFOOVYPCNVCYRGFOFMYSTPOPKYGAIFYFTKL  
KILAEKKIISAADKLIVFEAEVLITGFRCGHPIYIPPLSPRIDGHSYVDNS  
SPETVFIFSGQALIPYIMTCTOEVOSDYSSGPMRFAOHPMRGFASGSPVD"  
polyA\_signal  
1410..1415  
/gene="DKFZp761I1617"  
polyA\_site  
1431  
/gene="DKFZp761I1617"  
BASE COUNT 457 a 276 c 317 g 392 t  
ORIGIN  
Query Match 44.4%; Score 1440.4; DB 9; Length 1442;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1441: Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
polyA\_signal  
1410..1415  
/gene="DKFZp761I1617"  
polyA\_site  
1431  
/gene="DKFZp761I1617"  
BASE COUNT 457 a 276 c 317 g 392 t  
ORIGIN

Db 301 CAGTGAGCTATTTCAGACAAACAAAACCCAGAGCAAGTAAGAAATATCATTTCTG 360  
Qy 2161 AATGTCTCTGTCTCCCTCCAACTCAAGAGCTTCTAGATCAAAAACAGTTTAAAAATGT 2220  
Db 361 AATGTCTCTGTCTCCCTCCAACTCAAGAGCTTCTAGATCAAAAACAGTTTAAAAATGT 420  
Qy 2221 GGTTCGAGGCTCTAAGGTGAGAAATAGCAATGAGAGTCTTATGCTCCTTCA 2280  
Db 421 GGTTCGAGGCTCTAAGGTGAGAAATAGCAATGAGAGTCTTATGCTCCTTCA 480  
Qy 2281 AGAAGAAAGAAATATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2340  
Db 481 AGAAGAAAGAAATATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 540  
Qy 2341 CAGCAAGTCCCTTACACGTTCTGCAATGTGATGACAGAGTTGGCTTCAAGAAATGAC 2400  
Db 541 CAGCAAGTCCCTTACACGTTCTGCAATGTGATGACAGAGTTGGCTTCAAGAAATGAC 600  
Qy 2401 TCGACACCTTGGATCCAAATATAGGAGCTGGCATATACCTGACCAAGAACTCAAAAAC 2460  
Db 601 TCGACACCTTGGATCCAAATATAGGAGCTGGCATATACCTGACCAAGAACTCAAAAAC 660  
Qy 2461 CTGGCAGAGAAAGCCCAAGAAATCTCTGCTGAGATTAAGCTGATGTGTTGAGGCT 2520  
Db 661 CTGGCAGAGAAAGCCCAAGAAATCTCTGCTGAGATTAAGCTGATGTGTTGAGGCT 720  
Qy 2521 GAAGTACTCAGAGGCTTCTTCTGCCAGGACATCCGTTAAATATGTTCCTCCACACCTG 2580  
Db 721 GAAGTACTCAGAGGCTTCTTCTGCCAGGACATCCGTTAAATATGTTCCTCCACACCTG 780  
Qy 2581 ACTCTCGAGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640  
Db 781 ACTCTCGAGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Qy 2641 TTTGTTATTTTATAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700  
Db 841 TTTGTTATTTTATAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
Qy 2701 TATGTACATCAGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2760  
Db 901 TATGTACATCAGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 960  
Qy 2761 AGGGGATTCGCAAGTGGCAGCCCTGTTGATTAATGATTAATGATTAATGATTAATGATTAAT 2820  
Db 961 AGGGGATTCGCAAGTGGCAGCCCTGTTGATTAATGATTAATGATTAATGATTAATGATTAAT 1020  
Qy 2821 GGCCTTACCTTGGTGAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880  
Db 1021 GGCCTTACCTTGGTGAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080  
Qy 2881 ATATATCCCATGGGTTATCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940  
Db 1081 ATATATCCCATGGGTTATCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
Qy 2941 TAGCATCTTAGTCCCTTATCTGTCTTATGCTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT 3000  
Db 1141 TAGCATCTTAGTCCCTTATCTGTCTTATGCTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT 1200  
Qy 3001 TGAACACTTTCAGAGCTCTCTCTCTGAGAGTTGTTCTTAAATCTCTCTTACTAGAG 3060  
Db 1201 TGAACACTTTCAGAGCTCTCTCTCTGAGAGTTGTTCTTAAATCTCTCTTACTAGAG 3120  
Qy 3061 GAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120  
Db 1261 GAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
Qy 3121 TTTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3180  
Db 1321 TTTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
Qy 3181 AGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3240  
Db 1381 AGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440

QY 3241 AA 3242  
Db 1441 AA 1442

RESULT 6  
HSM804240  
LOCUS HSM804240 1388 bp mRNA linear PRI 10-JUL-2002  
DEFINITION Homo sapiens mRNA: cDNA DKFp666B0810 (from clone DKFp666B0810).  
ACCESSION AL832929  
VERSION AL832929.1 GI:21733516  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1388)  
Ansozge,W., Wlrkner,U., Mewes,H.W., Well,B. and Wiemann,S.  
Direct Submission  
Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by EMBL (European Molecular Biology Laboratories,  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
This clone (DKFp666B0810) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-charlottenburg, GERMANY; Email: clonerzpd.de Further  
information about the clone and the sequencing project is available  
at <http://mips.gsf.de/proj/cDNA/>.  
Location/Qualifiers  
source 1. 1388  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFp666B0810"  
/tissue\_type="stomach"  
/clone\_lib="666 (synonym: hsto2). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
polya\_signal 1342..1347  
polya\_site 1360  
BASE COUNT 450 a 257 c 302 g 379 t  
ORIGIN

Query Match 42.3%; Score 1371.8; DB 9; Length 1388;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1373; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1869 GAATATCATATTTCTGTACCTGGGAGAGAAAGACATGACATTTTGTCTCAGCTTCAGAA 1928  
Db 1 GAATATCATATTTCTGTACCTGGGAGAGAAAGACATGACATTTTGTCTCAGCTTCAGAA 60

QY 1929 AACTTCAAGTGTCTCCATCAGCAAAATTTATCAGCCCGAGAGCAAGATTAAGATTTGA 1988  
Db 61 AACTTCAAGTGTCTCCATCAGCAAAATTTATCAGCCCGAGAGCAAGATTAAGATTTGA 120

QY 1989 AGAGAGCCGGGCTGACCTGATTTGAGTGTATGACATTTGAAGATTTGCTTTAAAGT 2048  
Db 121 AGAGAGCCGGGCTGACCTGATTTGAGTGTATGACATTTGAAGATTTGCTTTTAAAGT 180

QY 2049 ACAGAGAGAAATGGCAAG 2108  
Db 181 ACAGAGAGAAATGGCAAG 240

QY 2109 TATTTCAGCAACAAACCAACCAAGACGAATGAAGAAATATCATATTTTCTGAAGTTC 2168  
Db 241 TATTTCAGCAACAAACCAACCAAGACGAATGAAGAAATATCATATTTTCTGAAGTTC 300

QY 2169 TGTGCGCTCAACCTCAAGAGCTGTAGATCAAAAGAAACAGTTTGAAGAAATGTGTTGCA 2228  
Db 301 TGTGCGCTCAACCTCAAGAGCTGTAGATCAAAAGAAACAGTTTGAAGAAATGTGTTGCA 360

QY 2229 GGTTCCTAAGGTGAGAGATAGACAAATGAGGTCTTATGGCTGCTTTCAAGAGAAAGAA 2288  
Db 361 GGTTCCTAAGGTGAGAGATAGACAAATGAGGTCTTATGGCTGCTTTCAAGAGAAAGAA 420

QY 2289 GAAATATGATGAGAGAAAGAAACCGCAGAGCAACCTGTAGCCATGAGGTGTTTCAGAGT 2348  
Db 421 GAAATATGATGAGAGAAAGAAACCGCAGAGCAACCTGTAGCCATGAGGTGTTTCAGAGT 480

QY 2349 CCCATACAGTTCGCAATGTGATGACAGAGTGTGGCTTTCAAGAAATGTACTGACACC 2408  
Db 481 CCCATACAGTTCGCAATGTGATGACAGAGTGTGGCTTTCAAGAAATGTACTGACACC 540

QY 2409 TTGCGATCCAAATACGAGAGCTGGCATATCTTCACCAAGACCTCAAAAACCTGGCAGA 2468  
Db 541 TTGCGATCCAAATACGAGAGCTGGCATATCTTCACCAAGACCTCAAAAACCTGGCAGA 600

QY 2469 GAAGGCCAAGAGAAATCTGCTGAGATTAAGCTGATATGTGTTTGAAGCTGAGTACT 2528  
Db 601 GAAGGCCAAGAGAAATCTGCTGAGATTAAGCTGATATGTGTTTGAAGCTGAGTACT 660

QY 2529 CACAGGCTTCCTGCGCAGGAGACATCCGTTAAATATTTGCCCCACACTGAGTCCCTGG 2588  
Db 661 CACAGGCTTCCTGCGCAGGAGACATCCGTTAAATATTTGCCCCACACTGAGTCCCTGG 720

QY 2589 AGCTATAGATGGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAACCTTTGTTAT 2648  
Db 721 AGCTATAGATGGTCATGACAGTGTGTTGACAAATGCTCCAGCCCTGAAACCTTTGTTAT 780

QY 2649 TTTTAGTGGCATGAGGCTTATCTCTAGTATTTGTGGACATGCCACCAAGAAATGTGACA 2708  
Db 781 TTTTAGTGGCATGAGGCTTATCTCTAGTATTTGTGGACATGCCACCAAGAAATGTGACA 840

QY 2709 GTCCAGAGATTACATCATCAGACACCAATGAGACCCCTTTGCACAGACATCTTGGAGGGATT 2768  
Db 841 GTCCAGAGATTACATCATCAGACACCAATGAGACCCCTTTGCACAGACATCTTGGAGGGATT 900

QY 2769 CGCAAGTGGGACGCCCTGTGATTATATCTCTACATCAATTTTAAACGCTGTATGCGCTTAC 2828  
Db 901 CGCAAGTGGGACGCCCTGTGATTATATCTCTACATCAATTTTAAACGCTGTATGCGCTTAC 960

QY 2829 CTTGGGTGAACCTAACCAATTAATGACATGATGGCTCAAGAGTGGCTTGAATATATTC 2888  
Db 961 CTTGGGTGAACCTAACCAATTAATGACATGATGGCTCAAGAGTGGCTTGAATATATTC 1020

QY 2889 CATGGTATATCTGTATGAGCTGACCTGAGTTATTTGAAGAGACTAGCCATATGACTCT 2948  
Db 1021 CATGGTATATCTGTATGAGCTGACCTGAGTTATTTGAAGAGACTAGCCATATGACTCT 1080

QY 2949 TAGTGCCTTATCTGTCTTATATGCTTGGGTTGGGTTAGGTAGATACCAATGAACAC 3008  
Db 1081 TAGTGCCTTATCTGTCTTATATGCTTGGGTTGGGTTAGGTAGATACCAATGAACAC 1140

QY 3009 TTTTCAGAGACCTTCCTGCTGAGTGTCTTAAATCCCTTACTAGAGGAGATAA 3068  
Db 1141 TTTTCAGAGACCTTCCTGCTGAGTGTCTTAAATCCCTTACTAGAGGAGATAA 1200

QY 3069 TATTTTCATATATGAAGAAATTTTCTAGTATATTAACGAGGCTTTTATTTTCTAA 3128  
Db 1201 TATTTTCATATATGAAGAAATTTTCTAGTATATTAACGAGGCTTTTATTTTCTAA 1260

QY 3129 ATGATGATAGTATTAATAATGTTAGAGATPAACAGATGATTTTACATTTTCCAGAGATAT 3188  
Db 1261 ATGATGATAGTATTAATAATGTTAGAGATPAACAGATGATTTTACATTTTCCAGAGATAT 1320

QY 3189 ATTAAGTCTTTAGTGTGAAGAAATATCATCTTGTGCTGATTAAGAAAGAAAGAA 3243  
Db 1321 ATTAAGTCTTTAGTGTGAAGAAATATCATCTTGTGCTGATTAAGAAAGAAAGAA 1375

RESULT 7  
BC003281 3012 bp mRNA linear ROD 07-AUG-2002  
LOCUS BC003281  
DEFINITION Mus musculus, clone MGC:7868 IMAGE:3581841, mRNA, complete cds.

ACCESSION BC003281  
VERSION BC003281.1 GI:13096972  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 3012)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amgdbcm.tmc.edu](mailto:amgdbcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>  
Series: IRAK Plate: 9 Row: 1 Column: 4  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13584917.  
Location/Qualifiers  
1. 3012  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="MGC:7868 IMAGE:3581841"  
/tissue\_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."  
/clone\_lib="NCI\_CGAP\_Mam1"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
122..2614  
/codon\_start=1  
/product="Unknown (protein for MGC:7868)"  
/protein\_id="AAH03281.1"  
/db\_xref="GI:13096973"  
/translation="NAVYMDTMAAARAPRANNSLEHYRMOPIKHNVEILKSNES QICEVLOKFKCISTISCTPLAGSSSPAPRVRRILPIELISWKDDLTRHVDVYV NANENLHLGSGIAGSLVKTGGEIDIESKRITIANQKISVGIATGTGRUPCHLI HAVGRWTVNSQTAIELKFAIRNLIDVTYDILIKTIVAPALSGLFOPDLCT STILETRILYFQDKMFGNLRHILNDSNPVAFKRSASESILGRDLSMGGPELDP ASMTLRIGRGLTLOVOCIEQMTDVTIVNSGYMDFSGVAOSILBOAGYEMKE LKVNISTDYOEVMWTKFRKLSQYVFHMAHAKSOLNKQIILDAKSCLEKLPKIN SISPRALGTGLMDLKKSTAAQLOIMEEVEVFAKEHKRITIVKIVIPVVEVETKIFIA EMKRSNELNLSGSLAOLWSSGEORRGLAEGPAINLKGKGECEQEWIER LVLVDHILIENNHILYLGKREHVDVLSLOTSTRVISSETVSPRTATLEIKPOADLI DAVMRLEDCICDVOEEVAGREKNLWLSGQCTNOGKDKMEESTFORYPASLTQI LODRRKQFEKCGILMYOVQIDNKKVLLAFORFKKKMERTRGSGSGLFOOVPHOE CMTVQNGPHRMKYSTXNPNVGATGYFTSTIKNLADKKYKTSSTDKLIVFPAEVLITG SFCQGNSSNIIIPPLSPGALDVNDSDVDVSSPEITIVNMQAMPLVIMTCTODRTE SOHPMWSQSYSSGPGVSSLSQSMEWLNSSV"

BASE COUNT 914 a 633 c 707 g 758 t  
ORIGIN

Query Match 42.0%; Score 1362.4; DB 10; Length 3012;  
Best Local Similarity 70.7%; Pred. No. 0;  
Matches 2065; Conservative 0; Mismatches 736; Indels 121; Gaps 14;

QY 367 CCTACAGTGTGCTGCTCTTGAGAAAATATAGTTGGCAATTCCTCCATT 426

Db 155 CCGCGGAAAGACACGCAACCAATTCCTTGAAGAATATATAGATGGCAATTCCTCATTT 214  
QY 427 AACCAATGACTTCAAAATTTTAAAAAATATAGACGCTGACGCTGTGTAAGTCTCCAG 486  
Db 215 AAACACAAATGTCTCAAAATTTTAAAGACCAATGACAGTCACTATGTAAGTCTCCAA 274  
QY 487 AATAAGTTGGGCTATCTACCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 543  
Db 275 AATAAGTTGGGCTATCTACCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 334  
QY 544 TCTCTCAAGTGTTCAGAAAATATGCTAGCTCCTAGAGTATGATGATGATGATGATGAT 603  
Db 335 GCTCAGAGAGTCTTCAAGAGGACCTGATCCCTGGGATGAGATGATGATGATGATGATGAT 394  
QY 604 GACCTACCAACATGCTCTGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 663  
Db 395 GACCTTACCAACATGCTCTGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 454  
QY 664 GGGGAGGCGCTGGCCCTGGCCCTGGTAAAAAGCTGGTAAATTTGAATCCAGAGAGAGC 723  
Db 455 GGAAGTGGCTGGCCGGAAGCTTGTGAAAACCTGTGTGAAATTCAGAGAGAGAGC 514  
QY 724 AAACAGTTTGTTCAGATATGTAAAGTGTGAGTGTGAGATGATGATGATGATGATGATGAT 783  
Db 515 AAAAGATATGATTCAGCAAGCTGTGTAATATCTGATGTGTGTGTGTGTGTGTGTGTGTGT 574  
QY 784 GGGAGGCTTCCCTGCAACAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 843  
Db 575 GGGAGGCTTCCCTGCAACAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 634  
QY 844 AAACAGGATATGCTGGAAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 903  
Db 635 AGCCAGACAGCTATGCAATTTACTGAAATTTGCCATTAGACATCTTACATTTATGTCACC 694  
QY 904 TATAAAATATCTACATTTAAGACAGTATGATTCACGCTTGAGTCTGTGGATTTTTCAG 963  
Db 695 AATATATGATCTAGCATTTAAGACAGTATGATTCACGCTTGAGTCTGTGGATTTTTCAG 754  
QY 964 TTCCTCTGATTTGTGTACAAAGACTATTTGTAGACATATCCGGTGTGTTGGCAAGGG 1023  
Db 755 TTCCCTCTGATTTGTGTACAAAGACTATTTGTAGACATATCCGGTGTGTTGGCAAGGC 814  
QY 1024 AAGCAGATATGATGATTTTGAAGAAATTTACCTGCTGTGAGCAATGAGGACCTTACTGTT 1083  
Db 815 AAGCAATGCTTGGTATTTTGAAGAAATTTACCTGCTGTGAGCAATGAGGACCTTACTGTT 874  
QY 1084 GCTGCTTTTAAAGCTCTTCAGATTCCTTGAAGAG-----AGTGAAGTGGGA 1134  
Db 875 GCTGCTTTTAAAGCTCTTCAGATTCCTTGAAGAGCTTGAAGAGCTTGAAGTGGGAGGT 934  
QY 1135 CAAGAAACCAAGCTTCTT-----TCAATGCAATGCTGTGAACAACCTGACCTCCAG 1188  
Db 935 CCAGAAACCAAGCTTCTT-----TCAATGCAATGCTGTGAACAACCTGACCTCCAG 994  
QY 1189 ATTGTCCAGGCGCACATTTGAATGACGAGCGAGATGATGATTTTGAATTTTGAATTTTGA 1248  
Db 995 ATTGTCCAGGCGCATTTGAATGACGAGCGAGATGATGATTTTGAATTTTGAATTTTGA 1054  
QY 1249 CATGATATTAACAGTTGACCTGTGGCAAGTCAATTTTACACAGCAGAGAGTTGAATG 1308  
Db 1055 CAGGATTTTAAATCAGAGCAGAGTGGCAGATGATCTTTGACAGAGAGGAGTTGAATG 1114  
QY 1309 AATTCGAAATTTCTTGGCACAAGGCTTAAACAGTTTCAACGAGGTCGATGCTGCTG 1368  
Db 1115 GAAAAGGAACTT-----GACAAGGTTAACCTGTCCACAGATTAATGAGAGGTGTGGTTC 1168  
QY 1369 ACAAGAGATTTTACCTGTTCTGTAATATATATATATATATATATATATATATATATAT 1428  
Db 1169 ACAAGAGATTTTAAATTTGCTCTGCAATGATGATGATGATGATGATGATGATGATGATG 1228  
QY 1429 CCTAAACCTCAGATATTTAAACATGCAATGCAAGAGTGTGTTGAAAAATTCATTTGAGCAA 1488  
Db 1429 CCTAAACCTCAGATATTTAAACATGCAATGCAAGAGTGTGTTGAAAAATTCATTTGAGCAA 1488

D	b	1229	AACAAATACCGATATTGGAAAGATGCATATGAGTCTCTGTCTGAAAAAAATGGCTTTAAACA	1288
O	y	1489	AATATACCTCCATTTCTCTTCCCTGCCCTTGGGACTGGAACATGGAAATAAAGAGAA	1548
D	b	1289	GATATTAATTCATTTCTCTTCTCTCGCTCGGGACGAGTATGCATTTTAAAGAGT	1348
O	y	1549	ACACAGCAGAGATTTTGTGTGATGAAGTTTAACTTTTCCAAAGACCATGTAAACAC	1608
D	b	1349	ACACAGCTCGATTAAGTTTGGAGGAAGTTTGGCATTTGGCTAAAGACACAGAAGAAA	1408
O	y	1609	CAGTTAATCTGTAATAATTTTGATCTTTCCACAGATTTTGGAACTATATATAGCTTTACGT	1668
D	b	1409	ACGCTAACTGTAAAGATTTGTGATCTTTCCAGTATGATGTGAGAACCTACAAAGATTTTAT	1468
O	y	1659	TCGTGAATGGCAAGAGGTCCAGATGCTGAGTTTGAACATTTACGT-----GTC	1719
D	b	1469	GCTGAATGACAAAAAGGTCCAAAGCGATGTAATCTCAGCGGTAAATAGTGTGCTTATGACC	1528
O	y	1720	CCCCAGTCACACAGAGAGGAAAAAGAAATGGGCTTTGAAGCTATGATCTCTGCCATC	1779
D	b	1529	CTGAGTGTGCTCACTGGGGAGCAAAAGAGGGGCTTTGAAGCTGTCTCCGCATC	1588
O	y	1780	AATCGATGGGATTCACAGCGGGAAGAGTATATGAGGGCCACGCTGGATCCAAAGAAATC	1839
D	b	1589	AATCTCATGGGTGTAAAAAGTGGAGAGATGTGTAGGCCCAAGATGGATTTGAAGGTTG	1648
O	y	1840	CTGAGTCTCCAGAACCCACCATCATTTGAGAAATATCATATTTCTTACCTTTGGGAGAA	1899
D	b	1649	CTGGTCTCCCTGGACCCACCATCATTTGAGAAATATCATATTTCTTATCTTTGGGAAAA	1708
O	y	1900	GACATGACATTTTGTCTCAGCTTCAGAAAACCTCAAGTGTCTCATCAGAAATATATC	1959
D	b	1709	GAGCAGCAGCGTGTCTGTGAGCTCCAGACACAGCAAGAGTCTTCATTTACAGACACTGC	1768
O	y	1960	AGCCAGGAAGAGACAGATGAGATTGGAAGGCGCCGGGCTGACCTCATTTAGAGTGGTT	2019
D	b	1769	AGTCCAAAGAAAGCGCCACTTTGGAGATTTAAAGTCCCCAGGCTGACTTATTTAGCAGATT	1828
O	y	2020	ATGAACATTTGAAGTATGCTTTGTAAAGTACAGAGAGAAATGGCAGGAAAAAGAGACGA	2079
D	b	1829	ATGAGATATGAATGTATGCTGTGTGACGTTCAGGAAGAAAGTGGCAGGAAAAAGGAGAA	1888
O	y	2080	GGCCTTTGGCGCTGTAGCAGAGTGGACTATTACAGACAAAAAACCAAGACGAATG	2139
D	b	1889	AATCTTTGGAGCTTGTACAGACAGGGGACCAACAGCAAGAAAAA---CTGATATAATG	1945
O	y	2140	AAACAAATATCATATTTTGTGAAATGCTGTGGCTCCACATCCAGAGCTCTATGATCAA	2199
D	b	1946	GAAATATCTACACATTTACAGCATCCACAGCATCTTATCTACAGGAACCTTCAGAGCCGA	2005
O	y	2200	AAGAAACAGTTTGAATAATGTGTTTGCAGAGTTCTTAAAGGTGGAGAAATAGACAAATGAG	2259
D	b	2006	AAGAAACAGTTTGAATAATGTGCTGTGTGGTTGTGAGGTGGACACATATGACAAATAG	2065
O	y	2260	GTCCTTATGCTGTGCTTTCAAGAGAAAGAAATGATGGAAGAAAACTGCACAGGCCAA	2319
D	b	2066	GTCGCTGTGCTGTCTTCACAGAGAGAGAAAGAAATGATGGAAGAGAGGCCCAAGAGGA	2125
O	y	2320	CCGTGGCCATAGCGTGTTCAGCAAGTCCCATACAGATTCGAAATGTGTATGACAGA	2379
D	b	2126	TCTGGGAGCAAAAGTTGGTTTCAGCGAGGTCCACATCAAGTTCTGCAATACGTTGTGAGA	2185
O	y	2380	GTTGGCTTTCAAGAAATGTACTGCAACCTTCGCATCCAAATACGAGAGTGGGATATAC	2439
D	b	2186	GTCGGCTTCCACAGAAATGTGTGACATCTCTATAACCCAGTTTATGAGAGCCGGCATATAT	2245
O	y	2440	TTTCAACAAGACCTCAAAAACCTCGCGAGAGAGGCCAAGAAATCTCTGCTCAGATATAG	2499
D	b	2246	TTTCAACAAGAGCTCAAAAATTTACGACAGACAAAGGTCAAGAAAACTCAAGCACAGACAG	2205
O	y	2500	CTGATCTATGTGTTTGAAGCTGAAGTACTACAGGCTTCTTCTGCGAGGAGCATCGCTTA	2559
D	b	2306	CTATCTATGTGTTTGAAGGAGAGATGATCTACAGAGGTCTCTTGTACAGAGTATATCTCTCA	2365

QY	2560	AAATATGTTCCCCACCACTAGTCCCTGGAGCTATAGATGTCATGACAGTGTGGTTGAC	2619
Db	2366	AAATATCATCCCTCCACACATTTGAGTCTCTGGGGCTTATAGTCAATGACAGCTAGTTGAC	2425
QY	2620	AATGTCCTCAGCCCTGAAAACCTTTGTATTTTATGAGCATCCAGGCTTACCTAGTAT	2679
Db	2426	AATGTTTCCAGCCCTGAAAACCAATGTGTGTTTTTAATGGCATCCAGGCCATGGCCCTGTAC	2485
QY	2680	TTGTGGACATGGACCCAGG-----AATATGTACAGTACCAAGATTAC	2721
Db	2486	TTGTGGACTTGGACACACAGATAGGACATTTCTACAGCATCCGATGTGGTACACAGGGCTAC	2545
QY	2722	TCATCAGGACCAATGAGACCCCTTTCACACGACATCCTTGGAGGGGATTTGCGCAAGTGGACAC	2781
Db	2546	TCATCAGGACCAAGGAATGGTCTTCCTCCGTCGACGTCTGTGGGAATGGGTCTTAAATGGCAGC	2605
QY	2782	CCTGTTGATTAATCTCTACATCATTTTAAACAGCTGGTATGCGCTTACCTTGGGGTAACTA	2841
Db	2606	TCTGT-----TTAAGTCTCATCATCAGTTTAAACAAGCAGAAAGG-----GTTCAGAGAACTGG	2656
QY	2842	ACCAAAATATGACCAATCCATGGCTCAAGAGTGGCTTGAAATATATCCATCCAGGCTATGCG	2901
Db	2657	ACAAAATGATTAATA-----ACAGGTTAACTCG	2683
QY	2902	TATGACTGACAGGGGTTATTTAAGAGACTAGCCACATATCTAGCATCTTAGTCCCTTATC	2961
Db	2684	TTTCAGATATGATGGGTCACTAAGAGCACCGACCGACACACTACATCATATAGTCCCTTGTGC	2743
QY	2962	TGTCATTATGTCCTTGGGGTTGGGGTAGGTAGATACCAATGAACACTTTCAGACCTTC	3021
Db	2744	-----TTTACTCTGGGCTTGTACGTGGGCGACATGCCACTTAAACTCTCTCCTCTCTT-	2796
QY	3022	CTTCCCTCTGCAGTGTGTTCTTAATCTCCCTTACTAGAGGAGATATAATTTTGGATATA	3081
Db	2797	TTCTATTTGACATCTTTATCTATCTCTTCCCTTAAGGTGACAGCAAGAATCTTATATATGA	2856
QY	3082	ATGAGAAATTTTCTAGATATATAAGCAGGCGCTTTATTTTCTTAAATATGATAGTATAT	3141
Db	2857	ACAAGGATATTTTTT-----CAAGCTGTATTTTCTTAAATGAT-----A	2898
QY	3142	AAAAATGTTAGATACAGAAATGATTTTATAGATTTCCAGAGAAATATATAAGTGCCTTAA	3201
Db	2899	GCACAAACTAGGACAACAAGATGATTTTCAAGGTTTCTATATATTTTATAAGTGCCTTGG	2958
QY	3202	GGTATGAAATTAATCATCTTGTCTGATTAATAAAAAAAAAA	3243
Db	2959	GATATCCAAATTAATCACCTTGTCTGTGATAAAAAAAAAA	3000

RESULT	8
LOCUS	AC026506
DEFINITION	167164 bp DNA linear HTG 11-APR-2000
ACCESSION	AC026506
VERSION	AC026506.2 GI:7534005
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 167164)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E. Homo sapiens chromosome 3, clone RP11-757I12 unpublished
JOURNAL	2 (bases 1 to 167164) Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,I., Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
REFERENCES	
AUTHORS	

TITLE  
JOURNAL  
COMMENT

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,R., Labouque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Leyne,R., Lieou,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,  
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tittell,A., Travers,M., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,W.

Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 11, 2000 this sequence version replaced g1:7283196.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L6563

Center clone name: 757-I\_12

Sequencing Statistics

Sequencing vector: M13: M77815: 100% of reads

Assembly program: Phrap, version 0.960731

Consensus quality: 143128 bases at least Q40

Consensus quality: 154634 bases at least Q20

Consensus quality: 159724 bases at least Q20

Insert size: 176000: agarose-fp

Quality coverage: 3.1 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-continfs

NOTE: This is a 'working draft' sequence. It currently  
consists of 33 continfs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the continfs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1059: contig of 1059 bp in length  
1060 1159: gap of 100 bp  
1160 3218: contig of 2059 bp in length  
3219 3318: gap of 100 bp  
3319 4939: contig of 1621 bp in length  
4940 5039: gap of 100 bp  
5040 5321: contig of 282 bp in length  
5322 5421: gap of 100 bp  
5422 6776: contig of 1355 bp in length  
6777 6876: gap of 100 bp  
6877 8440: contig of 1564 bp in length  
8441 8540: gap of 100 bp  
8541 10040: contig of 1500 bp in length  
10041 10140: gap of 100 bp  
10141 11902: contig of 1762 bp in length  
11903 12002: gap of 100 bp  
12003 15014: contig of 3012 bp in length  
15015 15114: gap of 100 bp  
15115 17618: contig of 2504 bp in length  
17619 17718: gap of 100 bp  
17719 20650: contig of 2932 bp in length  
20651 20750: gap of 100 bp  
20751 23090: contig of 2340 bp in length

23091 23190: gap of 100 bp  
23191 26671: contig of 3481 bp in length  
26672 26771: gap of 100 bp  
26772 29928: contig of 3157 bp in length  
29929 30028: gap of 100 bp  
30029 34433: contig of 4405 bp in length  
34434 34533: gap of 100 bp  
34534 39381: contig of 4848 bp in length  
39382 39481: gap of 100 bp  
39482 44572: contig of 5091 bp in length  
44573 44672: gap of 100 bp  
44673 49132: contig of 4460 bp in length  
49133 49232: gap of 100 bp  
49233 54475: contig of 5243 bp in length  
54476 54575: gap of 100 bp  
54576 58555: contig of 3880 bp in length  
58556 58655: gap of 100 bp  
58656 64694: contig of 6039 bp in length  
64695 64794: gap of 100 bp  
64795 70000: contig of 5206 bp in length  
70001 70100: gap of 100 bp  
70101 74717: contig of 4617 bp in length  
74718 74817: gap of 100 bp  
74818 81551: contig of 6734 bp in length  
81552 81651: gap of 100 bp  
81652 87788: contig of 6137 bp in length  
87789 87888: gap of 100 bp  
87889 93064: contig of 5176 bp in length  
93065 93164: gap of 100 bp  
93165 99571: contig of 6407 bp in length  
99572 99671: gap of 100 bp  
99672 108010: contig of 8339 bp in length  
108011 108110: gap of 100 bp  
108111 116297: contig of 8187 bp in length  
116298 116397: gap of 100 bp  
116398 126733: contig of 10336 bp in length  
126734 126833: gap of 100 bp  
126834 139797: contig of 12964 bp in length  
139798 139897: gap of 100 bp  
139898 153345: contig of 13448 bp in length  
153346 153445: gap of 100 bp  
15346 167164: contig of 13719 bp in length.  
167164 Location/Qualifiers  
1. 167164  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3"  
/clone="RP11-757112"  
/clone\_id="RPCT-11 Human Male BAC"  
1. 1059  
/note="assembly\_fragment"  
1160. 3218  
/note="assembly\_fragment"  
3319. 4939  
/note="assembly\_fragment"  
5040. 5321  
/note="assembly\_fragment"  
clone\_end:77  
vector\_side:left  
5422. 6776  
/note="assembly\_fragment"  
6877. 8440  
/note="assembly\_fragment"  
8541. 10040  
/note="assembly\_fragment"  
10141. 11902  
/note="assembly\_fragment"  
12003. 15014  
/note="assembly\_fragment"  
15115. 17618  
/note="assembly\_fragment"  
17719. 20650  
/note="assembly\_fragment"

misc\_feature

QY	920	TTAAGACGCTGCACATTCACGCTTGAGCTCCGGGATTTTTCAGTTCCCTCGAATTTGT	979		
Db	100224	TTAAGACGCTGCACATTCACGCTTGAGCTCCGGGATTTTTCAGTTCCCTCGAATTTGT	100283		
QY	980	GTACAAAGACTATGTTAGAGACTATCCGGGTTAGTTTCCAAAGGAAGCCAAATGATGAGTA	1039		
Db	100284	GTACAAAGACTATGTTAGAGACTATCCGGGTTAGTTTCCAAAGGAAGCCAAATGATGAGTA	100343		
QY	1040	ATTTAAAGAAATTCACCTGGTGAGCATGAGAGACCCCTACTGTCCTGCTTAAAGCTG	1099		
Db	100344	ATTTAAAGAAATTCACCTGGTGAGCATGAGAGACCCCTACTGTCCTGCTTAAAGCTG	100403		
QY	1100	CTTCGAAATTCATCTCAGGGAAGAGTGAAGCTGGGCAAGAAACACCCCTCTTCAATG	1159		
Db	100404	CTTCGAAATTCATCTCAGGGAAGAGTGAAGCTGGGCAAGAAACACCCCTCTTCAATG	100463		
QY	1160	CAATGTCGTGGAACACCTGCACCCCTCCAGATTGTCTCAGGGCCACATTTGATGGCAGACGG	1219		
Db	100464	CAATGTCGTGGAACACCTGCACCCCTCCAGATTGTCTCAGGGCCACATTTGATGGCAGACGG	100523		
QY	1220	CAGATGTAATTGTTAATTTCTGT	1241		
Db	100524	TAAAGCTTTGTTCTTAATTTCTGT	100545		
RESULT 9	AC092908	174257 bp	DNA	linear	PRI 16-MAY-2002
LOCUS	AC092908				
DEFINITION	Human sapiens 3q BAC RP11-57112 (Roswell Park Cancer Institute				
ACCESSION	AC092908				
VERSION	AC092908.9	GI:17223136			

RESULT	9	174257 bp	DNA	linear	PRI 16-MAY-2002
LOCUS	AC092908				
DEFINITION	Human sapiens 3q BAC RP11-757112 (Roswell Park Cancer Institute				
ACCESSION	AC092908	Human BAC Library) complete sequence.			
VERSION	AC092908.9	GI:1722136			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE  
1 (bases 1 to 174257)  
Muzny, D.M., Adams, C., Adio-Odola, B., All-  
osman, F.R., Allen, C.,  
Alstbrooks, S.L., Amaratunga, H.C., Aye, J.R.,  
Baks, T., Barbarta, J.,

Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R.,

Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Durrant, D., Edwards, J. G., Edwards, R., Edwards, C. C.

Hamilton, K., Harris, C., Harris, D., Hart, M., Haykal, D., Hayes, A., Elmhurst, J., Elmer, P., Escott, M., Fabbri, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Prantz, P., Gabisi, A., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, D., Hart, M., Haykal, D., Hayes, A.

He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Huiyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joubert, S., Karlsson, F., Kelly, S., Khan, H., Kind, I., Kovach, J.

[illegible]

Mallinckrodt, C., Massey, E., Mahoney, E., McCabe, M. P., Maatou, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newson, J., Newton, N., Nauven, A., Nauven, N., Nauven, N.

NICHOLSON, E., MWOOLUKWU, S., YOUNG, M., CAMERON, S., OJAGHIE, N., OVIDO, R., PACE, A., PAYTON, B., PEERY, J., PETER, L., PICKENS, R., PRIMUS, E., PU, L. L., QUILLS, M., REN, Y., RIVERS, M., ROJAS, A., ROTUPOKAN, I., ROLFE, M., RUIZ, S., SAVERY, G., SCHERER, S.,

Beckley, T., Chen, H., Brodsky, A., Cassan, A.,  
Sonakke, T., Sparks, A., Stanley, H., Stone, H.,  
Tabori, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tanton, J., Taylor, C., Sutton, A., Swatek, A.,  
Tansey, J., Taylor, C.,



```

/rpl_family="MERSA"
repeat_region      18997. .19025
/rpl_family="(CCCCC)n"
repeat_region      complement(20134. .20485)
/rpl_family="L2"
repeat_region      22275. .22626
/rpl_family="MER63C"
repeat_region      22627. .22650
/rpl_family="(TA)n"

Query Match      26.0%; Score 844.4; DB 9; Length 174257;
Best Local Similarity 98.7%; Pred. No. 3,7e-201;
Matches 851; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 380 CAGAGACTGCTGCTTGGAGAAACTATAGTTGGCAAAATCCCATTAACCAATGACT 439
|||||
Db 144092 CAGAGACTGCTGCTTGGAGAAACTATAGTTGGCAAAATCCCATTAACCAATGACT 144151

QY 440 TCAAAATTTTAAATATAGAGCGTCAGCTGTGTGAAGTCTCCAGATTAAGTTGACT 499
|||||
Db 144352 TCAAAATTTTAAATATAGAGCGTCAGCTGTGTGAAGTCTCCAGATTAAGTTGACT 144211

QY 500 GTATCTGACCCGTGCTCTCCAGTTGAGAGGCAACAGCAATCTCTGCAAGTTTCA 559
|||||
Db 144212 GTATCTGACCCGTGCTCTCCAGTTGAGAGGCAACAGCAATCTCTGCAAGTTTCA 144271

QY 560 GAAATATGCTGACTCTAGATAGATTATGCTGTGAAGATGACCTCCACACATG 619
|||||
Db 144272 GAAATATGCTGACTCTAGATAGATTATGCTGTGAAGATGACCTCCACACATG 144331

QY 620 CTGTGTGATGCTGTGTGAATGACCCCAATGAATCTTCTGATGGGGAGGCTGGCCC 679
|||||
Db 144332 CTGTGTGATGCTGTGTGAATGACCCCAATGAATCTTCTGATGGGGAGGCTGGCCC 144391

-QY 680 TGGCCCTGCTAAAGCTGTGTGATTTGAATTCACAAAGAGAGCAACAGTTTGTGCA 739
|||||
Db 144392 TGGCCCTGCTAAAGCTGTGTGATTTGAATTCACAAAGAGAGCAACAGTTTGTGCA 144451

QY 740 GATATGTGAAGTGTGAGTGTGAGATAGCTGACGAGGAGGAGGCTTCCCTGCA 799
|||||
Db 144452 GATATGTGAAGTGTGAGTGTGAGATAGCTGACGAGGAGGAGGCTTCCCTGCA 144511

QY 800 AACGATCATCATGCTGTGGGCTCGGTGATGGAATGGGATTAACAGGATGTACTG 859
|||||
Db 144512 AACGATCATCATGCTGTGGGCTCGGTGATGGAATGGGATTAACAGGATGTACTG 144571

QY 860 GAAAGCTGACAGAGGCCATTGTAATTTGCAATTTATGTATCTATATAAATCTACA 919
|||||
Db 144572 GAAAGCTGACAGAGGCCATTGTAATTTGCAATTTATGTATCTATATAAATCTACA 144631

QY 920 TTAAGACAGTATGTAAGATCTATCCGGGTTACTTTGCAAGGAGCAATGATGAGTA 1039
|||||
Db 144632 TTAAGACAGTATGTAAGATCTATCCGGGTTACTTTGCAAGGAGCAATGATGAGTA 144691

QY 980 GTACAAAGACTATTTGAGAGATATCCGGGTTACTTTGCAAGGAGCAATGATGAGTA 1039
|||||
Db 144692 GTACAAAGACTATTTGAGAGATATCCGGGTTACTTTGCAAGGAGCAATGATGAGTA 144751

QY 1040 ATTTGAAAGAAATTCACCTGTGTGAGCAATGAGACCTACTGTTGCTCTTTAAAGCTG 1099
|||||
Db 144752 ATTTGAAAGAAATTCACCTGTGTGAGCAATGAGACCTACTGTTGCTCTTTAAAGCTG 144811

QY 1100 CTTGAGAAATTCATCTGAGGAGAGTGAGTGAGGCAAAAGCAACCCCTTTCATG 1159
|||||
Db 144812 CTTGAGAAATTCATCTGAGGAGAGTGAGTGAGGCAAAAGCAACCCCTTTCATG 144871

QY 1160 CAATGCTGCTGAAACACTGACCTCCAGATTTTCAGAGGCCACATTTGAATGACGACG 1219
|||||
Db 144872 CAATGCTGCTGAAACACTGACCTCCAGATTTTCAGAGGCCACATTTGAATGACGACG 144931

QY 1220 CAGATGTAATTTGTTAAATTCGT 1241
|||||
Db 144932 TAACTCTTGTCTTCTATTCCT 144953

```

```

RESULT 10
AC026506/c
LOCUS DEFINITION
AC026506
Homo sapiens chromosome 3 clone RP11-757112 map 3, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
ACCESSION
AC026506.2 GI:7534005
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167164)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Homo sapiens chromosome 3, clone RP11-757112
Unpublished
2 (bases 1 to 167164)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgealter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,
McElrim,J., Menus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 11, 2000 this sequence version replaced gi:7283196.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6563
Center clone name: 757112
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 143128 bases at least Q40
Consensus quality: 154634 bases at least Q30
Consensus quality: 159724 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 163964; sum-of-ctnigs
Quality coverage: 3.1 in Q20 bases; sum-of-ctnigs
Quality coverage: 3.3 in Q20 bases; sum-of-ctnigs

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown.

```
* as soon as it is available and the accession number will
* be preserved.
1 1059: contig of 1059 bp in length
* 1060 1159: gap of 100 bp
* 1160 1218: contig of 2059 bp in length
* 3219 3318: gap of 100 bp
* 3319 4939: contig of 1621 bp in length
* 4940 5039: gap of 100 bp
* 5040 5321: contig of 282 bp in length
* 5322 5421: gap of 100 bp
* 5422 6776: contig of 1355 bp in length
* 6777 6876: gap of 100 bp
* 6877 8440: contig of 1564 bp in length
* 8441 8540: gap of 100 bp
* 8541 10040: contig of 1500 bp in length
* 10041 10140: gap of 100 bp
* 10141 11902: contig of 1762 bp in length
* 11903 12002: gap of 100 bp
* 12003 15014: contig of 3012 bp in length
* 15015 15114: gap of 100 bp
* 15115 17618: contig of 2504 bp in length
* 17619 17718: gap of 100 bp
* 17719 20650: contig of 2932 bp in length
* 20651 20750: gap of 100 bp
* 20751 23090: contig of 2340 bp in length
* 23091 23190: gap of 100 bp
* 23191 26671: contig of 3481 bp in length
* 26672 26771: gap of 100 bp
* 26772 29928: contig of 3157 bp in length
* 29929 30028: gap of 100 bp
* 30029 34433: contig of 4405 bp in length
* 34434 34533: gap of 100 bp
* 34534 39381: contig of 4848 bp in length
* 39382 39481: gap of 100 bp
* 39482 44572: contig of 5091 bp in length
* 44573 44672: gap of 100 bp
* 44673 49132: contig of 4460 bp in length
* 49133 49232: gap of 100 bp
* 49233 54475: contig of 5243 bp in length
* 54476 54575: gap of 100 bp
* 54576 58555: contig of 3960 bp in length
* 58556 58655: gap of 100 bp
* 58656 64694: contig of 6039 bp in length
* 64695 64794: gap of 100 bp
* 64795 70000: contig of 5206 bp in length
* 70001 70100: gap of 100 bp
* 70101 74717: contig of 4617 bp in length
* 74718 74817: gap of 100 bp
* 74818 81551: contig of 6734 bp in length
* 81552 81651: gap of 100 bp
* 81652 87788: contig of 6137 bp in length
* 87789 87888: gap of 100 bp
* 87889 93064: contig of 5176 bp in length
* 93065 93164: gap of 100 bp
* 93165 99571: contig of 6407 bp in length
* 99572 99671: gap of 100 bp
* 99672 108010: contig of 8339 bp in length
* 108011 108110: gap of 100 bp
* 108111 116297: contig of 8187 bp in length
* 116298 116397: gap of 100 bp
* 116398 126733: contig of 10336 bp in length
* 126734 126833: gap of 100 bp
* 126834 139797: contig of 12964 bp in length
* 139798 139897: gap of 100 bp
* 139898 153345: contig of 13448 bp in length
* 153346 153445: gap of 100 bp
* 153446 167164: contig of 13719 bp in length.
FEATURES
Source
1. 167164
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"

misc_feature
/clone="Rp11-757112"
/clone.lib="RPCr-11 Human Male BAC"
1. 1059
/note="assembly_fragment"
misc_feature
1160. 3218
/note="assembly_fragment"
misc_feature
3319. 4939
/note="assembly_fragment"
misc_feature
5040. 5321
/note="assembly_fragment"
/note_end:77
vector_side:left"
misc_feature
5422. 6776
vector_side:left"
misc_feature
6877. 8440
/note="assembly_fragment"
misc_feature
8541. 10040
/note="assembly_fragment"
misc_feature
10141. 11902
/note="assembly_fragment"
misc_feature
12003. 15014
/note="assembly_fragment"
15115. 17618
/note="assembly_fragment"
misc_feature
17719. 20650
/note="assembly_fragment"
20751. 23090
/note="assembly_fragment"
23191. 26671
/note="assembly_fragment"
26772. 29928
/note="assembly_fragment"
30029. 34433
/note="assembly_fragment"
34534. 39381
/note="assembly_fragment"
39482. 44572
/note="assembly_fragment"
44673. 49132
/note="assembly_fragment"
49233. 54475
/note="assembly_fragment"
54576. 58555
/note="assembly_fragment"
58656. 64694
/note="assembly_fragment"
64795. 70000
/note="assembly_fragment"
70101. 74717
/note="assembly_fragment"
74818. 81551
/note="assembly_fragment"
81652. 87788
/note="assembly_fragment"
87889. 93064
/note="assembly_fragment"
/note_end:sp6
vector_side:left"

Query Match 25.3%; Score 821; DB 2; Length 167164;
Best Local Similarity 100.0%; Pred. No. 2.9e+195;
Matches 821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2413 GATCCAAATACGACTGGCATATATCTTCACCAAGACCTCAAAAACCTGCAGAGANG 2472
DB 90675 GATCCAAATACGACTGGCATATATCTTCACCAAGACCTCAAAAACCTGCAGAGAG 90616
QY 2473 GCCAAGAAAATCTCTCTGAGATTAAGCTGATATGTTGAGGCTGAAGTACTCACA 2532
DB 90615 GCCAAGAAAATCTCTCTGAGATTAAGCTGATATGTTGAGGCTGAAGTACTCACA 90556
QY 2533 GGCCTTCTTGCAGGAGCATCCGTTAAATATTGTTCCCCACACCTGAGTCTGGAGCT 2592
```

Db 90555 GCGTCTTCTGCGCAGGAGACATCCGTTAAATATGTTCCCCACACATGAGTCTGAGCT 90496  
QY 2593 ATAGATGGTCATGACAGTGTGGTTGACAAATGTCTCAGGCCCTGAACCTTGTATTATTTT 2652  
Db 90495 ATATATGGTCAATGACAGTGTGGTTGACAAATGTCTCAGGCCCTGAACCTTGTATTATTTT 90436  
QY 2653 AGTGGCATGACAGGCTATACCTCACTATTTTGTGACATGACACCCAGGAATATGTACAGTCA 2712  
Db 90435 AGTGGCATGACAGGCTATACCTCACTATTTTGTGACATGACACCCAGGAATATGTACAGTCA 90376  
QY 2713 CAATATACATCAATGACAGCAATGACACCTTTCACAGACATCCTTGGAGGGGATTGGCA 2772  
Db 90375 CAATATACATCAATGACAGCAATGACACCTTTCACAGACATCCTTGGAGGGGATTGGCA 90316  
QY 2773 AGTGGCAGCCCTGTGATTAATCTCTACATCATTTTAAACAGCTGTATGGCCTTACCTTG 2832  
Db 90315 AGTGGCAGCCCTGTGATTAATCTCTACATCATTTTAAACAGCTGTATGGCCTTACCTTG 90256  
QY 2833 GGTGACCTAACCAATATATGACATGATGAGCTCAAGAGTGGCTTGAAATATATCCATG 2892  
Db 90255 GGTGACCTAACCAATATATGACATGATGAGCTCAAGAGTGGCTTGAAATATATCCATG 90196  
QY 2893 GGTATATCTGATGACATGACATGAGCTGTATGTAAGAGCTAGCACAATACATCTTATG 2952  
Db 90195 GGTATATCTGATGACATGACATGAGCTGTATGTAAGAGCTAGCACAATACATCTTATG 90136  
QY 2953 GCGTTATCTGCTCTTATATGCTTGGGGTGGGAGTGAATATCAATGAACACCTTTC 3012  
Db 90135 GCGTTATCTGCTCTTATATGCTTGGGGTGGGAGTGAATATCAATGAACACCTTTC 90076  
QY 3013 AGGACCTTCTCTCTCTGAGTGTCTTAACTCTCTTACTAGAGAGGATTAATATT 3072  
Db 90075 AGGACCTTCTCTCTCTGAGTGTCTTAACTCTCTTACTAGAGAGGATTAATATT 90016  
QY 3073 TTGATATATATGAAGAATTTTCTACTATATACGCAAGCCTTATTTTCTAAATGA 3132  
Db 90015 TTGATATATATGAAGAATTTTCTACTATATACGCAAGCCTTATTTTCTAAATGA 89956  
QY 3133 TGATAGATATATAAATGTTAGATATACGAATGATTTTTCAGAGAAATATTATA 3192  
Db 89955 TGATAGATATATAAATGTTAGATATACGAATGATTTTTCAGAGAAATATTATA 89896  
QY 3193 AGTCTTTAGGTATGAATAATCATCTTGTCTGATTA 3233  
Db 89895 AGTCTTTAGGTATGAATAATCATCTTGTCTGATTA 89855  
RESULT 11  
AX188088  
LOCUS AX188088 822 bp DNA linear PAT 06-AUG-2001  
DEFINITION Sequence 3783 from Patent WO0142467.  
ACCESSION AX188088  
VERSION AX188088.1 GI:15139560  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 822)  
Schlegel,R., Deeds,J., Berger,A. and Zhao,X.  
Genes, compositions, kits, and methods for identification,  
assessment, prevention, and therapy of cervical cancer  
Patent: WO 0142467-A 3783 14-JUN-2001;  
JOURNAL Millennium Predictive Medicine, Inc. (US)  
FEATURES  
source location/Qualifiers  
1..822  
BASE COUNT 228 a 177 c 211 g 198 t 8 others  
ORIGIN  
Query Match 19.28; Score 622.8; DB 6; Length 822;  
Best Local Similarity 86.1%; Pred. No. 1,4e-145;

Matches 759; Conservative 0; Mismatches 17; Indels 106; Gaps 2;  
QY 121 GCGGAGACGGCGGCGGAGAGTGGGCGACCATATCTGAAACATACAGTATCTTGA 180  
Db 45 GAGAGACGGCGCCCGGAGAGTGGGCGACCATATCTGAAACATACAGTATCTTGA 104  
QY 181 AGCGAAAGGGAATTAACATTTTAAAGATCCCGGGGACCTGGAGATGAGACTTTTC 240  
Db 105 AGCGAAAGGGAATTAACATTTTAAAGATCCCGGGGACCTGGAGATGAGACTTTTC 164  
QY 241 ATGTGGCCGGGAGAGAGCTTACATGAAATACAGTATGATTACTGCTCTCACTC 300  
Db 165 ATGTGGCCGGGAGAGAGCTTACATGAAATGAAAAA----- 197  
QY 301 TTGTTTCAAGAAAGCTTTCCTCAGATCTTTCCTCAGTGTGAGAAAGGGAATACAGAA 360  
Db 198 ----- 197  
QY 361 TGTCCTCCCTCAAGTCTCAGAGACTGTGCTCTTGGAGAAACATATAGTTGGCAATT 420  
Db 198 -----TCAGAGACTGTGCTCTTGGAGAAACATATAGTTGGCAATT 239  
QY 421 CCCATTAACCAATGACTTCAAAATTTTAAATAATATGAGCGTCAAGCTGTGAAGTC 480  
Db 240 CCCATTAACCAATGACTTCAAAATTTTAAATAATATGAGCGTCAAGCTGTGAAGTC 299  
QY 481 CTCAGATATAGTTTGGCTGTATCTTACCTGGTCTCCAGTTTCAGGAAGCAACAGC 540  
Db 300 CTCAGATATAGTTTGGCTGTATCTTACCTGGTCTCTCCAGTTTCAGGAAGCAACAGC 359  
QY 541 AAATCTCTGCAAGTGTTCAGAAAAATCTGACTCTAGATAGATAGATTATCAGTCTGAAA 600  
Db 360 AAATCTCTGCAAGTGTTCAGAAAAATCTGACTCTAGATAGATAGATTATCAGTCTGAAA 419  
QY 601 GATGACCTCACACACATCTGCTGTGATGCTGTGTGAATGACGCAATGAAGATCTTCTG 660  
Db 420 GATGACCTCACACACATCTGCTGTGATGCTGTGTGAATGACGCAATGAAGATCTTCTG 479  
QY 661 CATGGGGGAGGCGCTGGCCCTGGCCCTGTGTAAGACCTGTGTAATTCGAATTCGAAGAAG 720  
Db 480 CATGGGGGAGGCGCTGGCCCTGGCCCTGTGTAAGACCTGTGTAATTCGAATTCGAAGAAG 539  
QY 721 AGCAAAAGTGTGTGCGCAGATATGTAAGTGTGACCTGGTGAATAGCTGTACGCGA 780  
Db 540 AGCAAAAGTGTGTGCGCAGATATGTAAGTGTGACCTGGTGAATAGCTGTACGCGA 599  
QY 781 GCAAGGAGGCTCTCCCTGCAAAACAGATCATCTGTTGGCCCTCGGTGATGGAATGG 840  
Db 600 GCAAGGAGGCTCTCCCTGCAAAACAGATCATCTGTTGGCCCTCGGTGATGGAATGG 659  
QY 841 GATTAACAGGAGTGTGGAAGAGCTGACAGAGG-CCATTGTAGTATTTCTGAATTATGT 899  
Db 660 GATTAACAGGAGTGTGGAAGAGCTGACAGAGGCGCATGTGTAGTATTCGGAAATATGT 719  
QY 900 CATCTATATAAATATCTACATTAAGACAGTATTCAGCCTTGGAGCTTGGGATTTT 959  
Db 720 CATCTATATAAATATCTACATTAAGACAGTATTCAGCCTTGGAGCTTGGGATTTT 779  
QY 960 TCAGTTCCTCTGATTTGTTGTTGTTCAAGACTATTTGTGAAGAC 1001  
Db 780 TCAGTTCCTCTGATTTGTTGTTGTTCAAGACTATTTGTGAAGAC 821  
RESULT 12  
AX188241  
LOCUS AX188241 666 bp DNA linear PAT 06-AUG-2001  
DEFINITION Sequence 3936 from Patent WO0142467.  
ACCESSION AX188241  
VERSION AX188241.1 GI:15139714  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	Schlegel, R., Deeds, J., Berger, A. and Zhao, X.	Genes, compositions, kits, and methods for identification, assessment, prevention, and therapy of cervical cancer	Patent: WO 014267-A 3936 14-JUN-2001;	Millennium Predictive Medicine, Inc. (US)
1 (bases 1 to 666)				Location/Qualifiers 1..666
BASE COUNT	191 a	139 c	183 g	152 t 1 others
ORIGIN	/db_xref="taxon:9606"			
Query Match	Best Local Similarity	84.6%; Pred. No. 5.4e-112;	Matches 608; Conservative 0; Mismatches 16; Indels 105; Gaps 1;	
OY	136 GCAGAAATGGGCGACCATATCTGGAAGTACAGTGTATGCTTTGAAGGCCAAGAAGGAT	195		
DB	52 GAAAAAGTGCCGCACCATTATCTGAAACTTACAGTCTTAAGCTTTGAAGGCCAAGAAGGAT	111		
OY	196 AAACAATTTAAGACATCCCCGGGACCTGGAAGATGACATTTTCATGGTGGCCGAGACA	255		
DB	112 AAACAATTTAAGACATCCCCGGGACCTGGAAGATGACATTTTCATGGTGGCCGAGACA	171		
OY	256 GCAGCTTCAATGAAAATCAGTAGATGATTTACCTGCTCTCAGTCTGTTTCAGAAAGTC	315		
DB	172 GCAGCTTCAATGAAAATCAGTAGATGATTTACCTGCTCTCAGTCTGTTTCAGAAAGTC	189		
OY	316 TTTGCTCAGATCTTCTCAGTAGGAGAAAGGGGATACAGAAGATGTCTCCCTAACAG	375		
DB	190 -----	189		
OY	376 TGCTCAGAGACGTGCTCTTTGAGAAAACATATAGTTGGCAATTCATTAACCAAT	435		
DB	190 ---TCAGAGACGTGCTCTTTGAGAAAACATATAGTTGGCAATTCATTAACCAAT	246		
OY	436 GACTTCAAAATTTTAAATAATATAGCGCTCAGCTGTGTGAAGTCTCCAGATTAAGTTT	495		
DB	247 GACTTCAAAATTTTAAATAATATAGCGCTCAGCTGTGTGAAGTCTCCAGATTAAGTTT	306		
OY	496 GGCCTGATCTCACCTGGTCTCCAGTCCAGTCCGAAAGCAACAGCAATCTCTGCAAGT	555		
DB	307 GGCCTGATCTCACCTGGTCTCTCAGTCCGAAAGCAACAGCAATCTCTGCAAGT	366		
OY	556 TTCAAAGAAAATCTGACTCTCTAGATAGATATAGTCTGGAAGATGACTCCACACA	615		
DB	367 TTCAAAGAAAATCTGACTCTCTAGATAGATATAGTCTGGAAGATGACTCCACACA	426		
OY	616 CATGCTGTTGATGCTGTGGTGAATCAGCCCAATGAAGATCTTCTCATGGGGAGGCGTG	675		
DB	427 CATGCTGTTGATGCTGTGGTGAATCAGCCCAATGAAGATCTTCTCATGGGGAGGCGTG	486		
OY	676 GCCCTGGCCCTGGTAAAGCTGTGTGATTTGAATTCAGAAAGACCAACAGATTGTT	735		
DB	487 GCCCTGGCCCTGGTAAAGCTGTGTGATTTGAATTCAGAAAGACCAACAGATTGTT	546		
OY	736 GCCAATATGGAAGTGTGACGCTGTGATAGCTGTACGCGGAGACAGGAGGCTTCCC	795		
DB	547 GCCAATATGGAAGTGTGACGCTGTGATAGCTGTACGCGGAGACAGGAGGCTTCCC	606		
OY	796 TGCAACAGATCATCATGCTGTTGGGCTCGGTGATGGAATGGAATTAACAGGATG	854		
DB	607 TGCAACAGATCATCATGCTGTTGGGCTCGGTGATGGAATGGAATTAACAGGATG	665		
RESULT 13	AXI85819/c	516 bp	DNA	Linear PAT 06-AUG-2001
LOCUS	AXI85819	Sequence 1514 from Patent W00142467.		
DEFINITION	AXI85819			
ACCESSION	AXI85819.1	GI:15137230		
VERSION				

KEYWORDS	SOURCE
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eumalypota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 516)
FEATURES	Schlegel,R., Deeds,J., Berger,A. and Zhao,X. Genes, compositions, kits, and methods for identification, assessment, prevention, and therapy of cervical cancer Patent: WO 0142467-A 1514 14-JUN-2001; Millennium Predictive Medicine, Inc. (US) location/Qualifiers
source	1..516 /organism='Homo sapiens' /db_xref='taxon:9606'
BASE COUNT	124 a 138 c 108 g 146 t
ORIGIN	
Query Match	14.3% Score 463.2; DB 6; Length 516;
Best Local Similarity	99.2%; Pred.No.2.1e-105;
Matches	476; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
OY	379 TCAGAGACTGTGTCCTTGGAGAAACATATAGTTGGCAA--TTCCCATTAACCAATATGA 437
DB	507 TCAGAGATTGGTGCTTGGAGAAACATATAGTTGGCAAATTTCCCATTAACCAATATGA 448
OY	438 CTTCAAAATTTTTAAAAAATATAGCGCTGCAGCTGTGTGAAGTCTCTCAGAAATAGTTGG 497
DB	447 CTTCAAAATTTTTAAAAAATATAGCGCTGCAGCTGTGTGAAGTCTCTCAGAAATAGTTGG 388
OY	498 CTGTATCTCTACCCTGGTCTCCAGTTCCAGGAAGCAACAGCAATCTGCAAGTGT 557
DB	387 CTGTATCTCTACCCTGGTCTCCAGTTCCAGGAAGCAACAGCAATCTGCAAGTGT 328
OY	558 CAGAAAATATGCTGACCTCTAGAGATAGAGTTATCACTGGAAGAATGACCTCACACACA 617
DB	327 CAGAAAATATGCTGACCTCTAGAGATAGAGTTATCACTGGAAGAATGACCTCACACACA 268
OY	618 TGCTGTGATGCTGTGTGATGACGCCAATGAAAGATCTTCTGCATGGGGAGGCTTGGC 677
DB	267 TGCTGTGATGCTGTGTGATGACGCCAATGAAAGATCTTCTGCATGGGGAGGCTTGGC 208
OY	678 CCTGGCCCTGTGTAAGCTGTGTGATTTTGAATCCAAAGAGAGCAAAACATTTGTTGC 737
DB	207 CCTGGCCCTGTGTAAGCTGTGTGATTTTGAATCCAAAGAGAGCAAAACATTTGTTGC 148
OY	738 CAGATATGTTAAAGTGTGACGTGGAGATAGCTCTACAGGGAGAGGAGGCTTCCCTG 797
DB	147 CAGATATGTTAAAGTGTGACGTGGAGATAGCTCTACAGGGAGAGGAGGCTTCCCTG 88
OY	798 CAACAGATCATCATGCTGTGGGCTTCGGTGGATGGGATGGAATAAACAGAGATGTAC 857
DB	87 CAACAGATCATCATGCTGTGGGCTTCGGTGGATGGGATGGAATAAACAGAGATGTAC 28
RESULT 14	
LOCUS	AX187250 481 bp DNA linear PAT 06-AUG-2001
DEFINITION	Sequence 2945 from Patent W00142467.
ACCESSION	AX187250
VERSION	AX187250.1 GI:15138702
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eumalypota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 481)
JOURNAL	Schlegel,R., Deeds,J., Berger,A. and Zhao,X. Genes, compositions, kits, and methods for identification, assessment, prevention, and therapy of cervical cancer Patent: WO 0142467-A 2945 14-JUN-2001; Millennium Predictive Medicine, Inc. (US) Location/Qualifiers

```
source 1. .481
/organism="Homo sapiens"
/db.xref="taxon:9606"
BASE COUNT 116 a 129 c 102 g 134 t
ORIGIN

Query Match 13.9%; Score 449.8; DB 6; Length 481;
Best Local Similarity 99.4%; Pred. No. 4.8e-102;
Matches 462; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 394 CTTGGAGAAACATATAGTTGGCAA--TTCCATTAAACCAATAGCTTCAAAATTTTAA 452
DB CTTGGAGAAACATATAGTTGGCAA--TTCCATTAAACCAATAGCTTCAAAATTTTAA 422
QY 481 CTTGGAGAAACATATAGTTGGCAA--TTCCATTAAACCAATAGCTTCAAAATTTTAA 422
DB CTTGGAGAAACATATAGTTGGCAA--TTCCATTAAACCAATAGCTTCAAAATTTTAA 422
QY 453 AAATATAGACCTGACGTGTGAAGCTCCAGAAATAGTTGGCTGTATCTACCT 512
DB AAATATAGACCTGACGTGTGAAGCTCCAGAAATAGTTGGCTGTATCTACCT 362
QY 421 CAATATAGACCTGACGTGTGAAGCTCCAGAAATAGTTGGCTGTATCTACCT 362
DB CAATATAGACCTGACGTGTGAAGCTCCAGAAATAGTTGGCTGTATCTACCT 302
QY 513 GGTCTCTCCAGTTGAGAGGCAACAGCAATCTGCAAGTTCAGAAAAATGCTGAC 572
DB GGTCTCTCCAGTTGAGAGGCAACAGCAATCTGCAAGTTCAGAAAAATGCTGAC 302
QY 573 TCTAGATAGATTTATCACTCTGAAAGATGACCTCACACATGCTGTATGCTGT 632
DB TCTAGATAGATTTATCACTCTGAAAGATGACCTCACACATGCTGTATGCTGT 242
QY 301 TCTAGATAGATTTATCACTCTGAAAGATGACCTCACACATGCTGTATGCTGT 242
DB TCTAGATAGATTTATCACTCTGAAAGATGACCTCACACATGCTGTATGCTGT 182
QY 633 GGTGAATGACCCAAATGAAGATCTTCATGAGGGAGGCTGGCCCTGGCTGTAA 692
DB GGTGAATGACCCAAATGAAGATCTTCATGAGGGAGGCTGGCCCTGGCTGTAA 182
QY 241 GGTGAATGACCCAAATGAAGATCTTCATGAGGGAGGCTGGCCCTGGCTGTAA 182
DB GGTGAATGACCCAAATGAAGATCTTCATGAGGGAGGCTGGCCCTGGCTGTAA 182
QY 693 AGCTGGTGGATTTGAATCCAGAGAGCAACAGTTGTTGCCAGATATGTAAAGT 752
DB AGCTGGTGGATTTGAATCCAGAGAGCAACAGTTGTTGCCAGATATGTAAAGT 122
QY 181 AGCTGGTGGATTTGAATCCAGAGAGCAACAGTTGTTGCCAGATATGTAAAGT 122
DB AGCTGGTGGATTTGAATCCAGAGAGCAACAGTTGTTGCCAGATATGTAAAGT 122
QY 753 GTCAGCTGGATATGCTGTCAAGGAGGAGGCTCCCTGCAACAGATCATCA 812
DB GTCAGCTGGATATGCTGTCAAGGAGGAGGCTCCCTGCAACAGATCATCA 62
QY 121 GTCAGCTGGATATGCTGTCAAGGAGGAGGCTCCCTGCAACAGATCATCA 62
DB GTCAGCTGGATATGCTGTCAAGGAGGAGGCTCCCTGCAACAGATCATCA 62
QY 813 TGCTGTTGGGCTCGGTGATGATGCAATGCAATGAGGATGTAC 857
DB TGCTGTTGGGCTCGGTGATGATGCAATGCAATGAGGATGTAC 17

RESULT 15
AX184391 457 bp DNA linear PAT 06-AUG-2001
LOCUS AX184391
DEFINITION Sequence 86 from Patent W00142467.
ACCESSION AX184391
VERSION AX184391.1 GI:15135742
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 457)
AUTHORS Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
TITLE Genes, compositions, kits, and methods for identification,
JOURNAL assessment, prevention, and therapy of cervical cancer
Patent: WO 0142467-A 86 14-JUN-2001;
FEATURES
Millennium Predictive Medicine, Inc. (US)
source 1. .457
Location/Qualifiers
BASE COUNT 108 a 126 c 98 g 125 t
ORIGIN

Query Match 12.8%; Score 414.8; DB 6; Length 457;
Best Local Similarity 98.2%; Pred. No. 3.1e-93;
Matches 430; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 421 CCATTAACCAATGACTTC-AAAATTTTAAAAATATAGCGTCAAGCTGTGTGAAGT 479
DB CCATTAACCAATGACTTC-AAAATTTTAAAAATATAGCGTCAAGCTGTGTGAAGT 398
QY 480 CCTCCAGAAATAGTTGGCTGTATCTCTACCTGCTCTCCAGTTCCAGAGGAAACAG 539
DB CCTCCAGAAATAGTTGGCTGTATCTCTACCTGCTCTCCAGTTCCAGAGGAAACAG 338
QY 397 CCTCCAGAAATAGTTGGCTGTATCTCTACCTGCTCTCCAGTTCCAGAGGAAACAG 539
DB CCTCCAGAAATAGTTGGCTGTATCTCTACCTGCTCTCCAGTTCCAGAGGAAACAG 278
QY 540 CAAATCTGCAAGTGTTCAGAAAAATGCTGACTCTAGAGTATGATATCACTGAA 599
DB CAAATCTGCAAGTGTTCAGAAAAATGCTGACTCTAGAGTATGATATCACTGAA 218
QY 600 AGATGACCTTACACACATGCTGTATGCTGTGTGTAATGCAAGCAATGAATCTTC 659
DB AGATGACCTTACACACATGCTGTATGCTGTGTGTAATGCAAGCAATGAATCTTC 218
QY 277 AGATGACCTTACACACATGCTGTATGCTGTGTGTAATGCAAGCAATGAATCTTC 659
DB AGATGACCTTACACACATGCTGTATGCTGTGTGTAATGCAAGCAATGAATCTTC 158
QY 660 GCATGGGGAGGCTGGCCCTGGCTGTTAAAGCTGTGATGTTGAAATCCAAAGA 719
DB GCATGGGGAGGCTGGCCCTGGCTGTTAAAGCTGTGATGTTGAAATCCAAAGA 158
QY 217 GCATGGGGAGGCTGGCCCTGGCTGTTAAAGCTGTGATGTTGAAATCCAAAGA 719
DB GCATGGGGAGGCTGGCCCTGGCTGTTAAAGCTGTGATGTTGAAATCCAAAGA 98
QY 720 GAGCAAAACAGTTTGTTCAGATATGTAAAGTGTGAGTGTGATGCTGACGGG 779
DB GAGCAAAACAGTTTGTTCAGATATGTAAAGTGTGAGTGTGATGCTGACGGG 98
QY 157 GAGCAAAACAGTTTGTTCAGATATGTAAAGTGTGAGTGTGATGCTGACGGG 779
DB GAGCAAAACAGTTTGTTCAGATATGTAAAGTGTGAGTGTGATGCTGACGGG 98
QY 780 AGCAGGAGGCTTCCTGCAACAGATCATCATGCTGTTGGGCTCGGTGATGGAATG 839
DB AGCAGGAGGCTTCCTGCAACAGATCATCATGCTGTTGGGCTCGGTGATGGAATG 38
QY 97 AGCAGGAGGCTTCCTGCAACAGATCATCATGCTGTTGGGCTCGGTGATGGAATG 839
DB AGCAGGAGGCTTCCTGCAACAGATCATCATGCTGTTGGGCTCGGTGATGGAATG 20
QY 840 GGATTAACAGGATGTAC 857
DB GGATTAACAGGATGTAC 20
```

```
Search completed: February 7, 2003, 04:00:51
Job time : 6738. secs
```